

Portland State University

PDXScholar

---

Dissertations and Theses

Dissertations and Theses

---

Summer 8-11-2017

# Evolution and Metabolic Potential of *Francisella*-like Endosymbionts of Ticks

Jonathan Graham Gerhart  
*Portland State University*

Follow this and additional works at: [https://pdxscholar.library.pdx.edu/open\\_access\\_etds](https://pdxscholar.library.pdx.edu/open_access_etds)



Part of the [Biology Commons](#)

Let us know how access to this document benefits you.

---

## Recommended Citation

Gerhart, Jonathan Graham, "Evolution and Metabolic Potential of *Francisella*-like Endosymbionts of Ticks" (2017). *Dissertations and Theses*. Paper 3832.  
<https://doi.org/10.15760/etd.5726>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: [pdxscholar@pdx.edu](mailto:pdxscholar@pdx.edu).

Evolution and Metabolic Potential of *Francisella*-like Endosymbionts of Ticks

by

Jonathan Graham Gerhart

A thesis submitted in partial fulfillment of the  
requirements for the degree of

Master of Science  
in  
Biology

Thesis Committee:  
Rahul Raghavan, Chair  
Anna-Louise Reysenbach  
Susan Masta

Portland State University  
2017

## ABSTRACT

Endosymbiosis in arthropods involves intracellular bacteria that supply an array of benefits to the host. Endosymbionts likely enhance the health of ticks by provisioning amino acids such as cysteine and tyrosine, and cofactors such as biotin and folic acid that are not available in blood – the sole nutrient source of ticks. Endosymbionts of ticks are of special interest due to their close evolutionary relationship with tick-vectoring pathogens that impact livestock and human health. For example, ticks typically contain *Coxiella*-like endosymbionts (CLEs) that are the closest relatives of the human pathogen *Coxiella burnetii*. In order to understand the evolutionary relationship between the mammalian pathogen *Francisella tularensis*, which is vectored by ticks, and the *Francisella*-like endosymbionts (FLEs) present in several ticks, we assembled the genomes of the FLEs in the hard tick *Amblyomma maculatum* and the soft tick *Ornithodoros moubata* using high-throughput sequencing. While this project was in progress, another group described the genome of an FLE in the soft tick *Argas (Persicargas) arboreus*. Utilizing the three genomes, we show that all FLEs evolved from a mammalian pathogen, a relationship that is converse to that of *C. burnetii*, which likely evolved from a tick-associated non-pathogenic ancestor. Additionally, our analyses indicate that FLEs are horizontally transferred between ticks, and due to their superior metabolic capabilities could replace ancestral endosymbionts with reduced genomes.

## Table of Contents

Abstract.....	i
List of Tables .....	iii
List of Figures.....	iv
Introduction.....	1
Methods.....	5
Results.....	8
Discussion.....	17
References.....	22
Appendix A	
Pseudogene analysis.....	30
Appendix B	
Associated Publications.....	31

## List of Tables

Table 1: Genome Features of FLEs .....	9
Table 2: Virulence genes in FLEs compared to <i>F. tularensis</i> .....	16

## List of Figures

Figure 1: Comparison of the FLE genomes.....	10
Figure 2: Comparison of metabolic features of FLEs.....	12
Figure 3: Phylogeny of FLEs.....	14
Figure 4: Comparison of FLE and host phylogenies.....	15
Figure 5: Comparison of metabolic potential of FLE-Am and CLEAA.....	18

## INTRODUCTION

Symbiotic relationships with bacteria have been observed in all domains of life, including intestinal symbionts that aid animals in digestion of complex foods, endophytic bacteria that support plants and fungi, and specialized arrangements such as bioluminescence in squid photophores (Moreira & López-García 1998; Hooper 2001; Nyholm & McFall-Ngai 2004). In arthropods, endosymbiosis, a specialized form of intracellular symbiosis, is observed where bacteria persist in the cytoplasm of host cells and provide a wide range of adaptive traits to the hosts, including nitrogen fixation, synthesis of complex vitamins and cofactors, provisioning of amino acids, and immune defense (Baumann 2005; Sabree et al. 2009; Weiss et al. 2011; Hansen & Moran 2011). Because the endosymbionts are sequestered within the rich media of the cytoplasm in comparatively low density, the lack of competition coupled with the access to abundant resources available within the host cell lead to a loss of selective pressure to retain superfluous genes (Moran 1996; Dale 2003). This downward spiral of genome decay is an “evolutionary rabbit hole” because the endosymbiont is sequestered in an environment that it cannot escape due to gene loss, and will continue to lose genes as it remains, leading to further dependency on the host. This decay will eventually lead to the loss of increasingly significant elements that support the host, eventually driving the host to replace the symbiont either through the acquisition of a new endosymbiont, or the supplementation of a reduced symbiont with a second endosymbiont (Bright & Bulgheresi 2010; Bennett & Moran 2015).

Among arthropods, endosymbionts of Ticks (order Ixodida) are an area of increasing interest. Ticks are organized into two major families: (1) Ixodidae, or hard ticks, which possesses a hardened chitinous scutum that anchors the head and serves as a

protective structure prior to feeding, and (2) Argasidae, or soft ticks, which do not have a rigid scutum, instead have mouthparts located on the underside of the body, and a thick, leathery coating that protects the tick. Hard ticks are thought to have arisen between 90 and 120 million years ago, as a derivative of arachnid insectivores that hunted prey using a set of spear-like mouthparts. Soft ticks are thought to have diverged from hard ticks approximately 30 million years ago, trading the armored scutum and segmented legs for greater mobility and blood capacity (Nava et al. 2009).

Ticks depend on an imbalanced diet consisting entirely of blood, which lacks several vital nutrients (Noda et al. 1997; Smith et al. 2015). Hence, tick endosymbionts are thought to ensure the health of the host through the synthesis of amino acids such as cysteine, the metabolism of highly concentrated amino acids such as serine and the synthesis of cofactors such as biotin (Smith et al. 2015). Unlike most primary insect endosymbionts, tick endosymbionts do not persist within specialized compartments called bacteriocytes (e.g. as observed in aphids), or among major structures of the carapace as seen in weevils (Baumann 2005; Lamelas et al. 2011). Instead, endosymbionts of ticks can be found in a large number of tissues, most importantly in the midgut tissues and in the salivary glands (Klyachko et al. 2007; Budachetri et al. 2014). Critically, the colonization of salivary glands is a common vector stage for bacterial pathogens borne by ticks, and saliva introduced to the host during feeding is the primary method of transmission of tick borne pathogens such as *Borriella* and *Rickettsia* (Spielman et al. 1987; Wright et al. 2015).



All ticks sequester endosymbionts mostly from the genera *Coxiella*, *Rickettsia* or *Francisella*. In addition to these three primary endosymbionts (i.e. necessary for host survival), secondary symbionts that are not strictly necessary for survival are found in varying prevalence throughout both hard and soft ticks, including *Midichloria*, *Arsenophonus*, *Rickettsiella*, *Wolbachia*, *Spiroplasma*, *Cardinium* and *Lariskella* species (Black & Piesman 1994; Klompen 2000; Ahantarig et al. 2013; Duron et al. 2017).

Endosymbionts of ticks are of special interest due to the agricultural impact of ticks through the destruction of cattle and other livestock (Jongejan & Uilenberg 2005; Fitzpatrick 2013), and the medical impact of their vectored diseases (Randolph et al. 1996). Additionally, the three primary endosymbionts of ticks share close evolutionary relationships with tick-borne pathogens *Rickettsia parkeri*, *Coxiella burnetii* and *Francisella tularensis*, making their investigation potentially medically significant (Ahantarig et al. 2013). CLEs are a paraphyletic group that is divided into three clades, each containing representatives from multiple genera of hard and soft ticks (Duron et al. 2014). All pathogenic *Coxiella*, such as *C. burnetii*, are clustered within one clade of endosymbionts, with only a few CLEs sharing a recent common ancestor with *C. burnetii*, suggesting that endosymbiosis predated the pathogenic phenotype (Smith et al. 2015). Because of this, it has been proposed that *C. burnetii* potentially arose from a tick symbiont (Duron et al. 2015). The other two primary tick endosymbionts, *Rickettsia*- and *Francisella*-like endosymbionts (FLEs) are thought to have undergone multiple transitions between symbiosis and pathogenesis (Duron et al. 2017).

FLEs are the third most common endosymbionts described in ticks, after *Coxiella* and *Rickettsia* (Scoles 2004; Budachetri et al. 2014). It is important that we understand the evolution of this clade of endosymbionts in more detail because of the symbiont's potentially close relationship to the human pathogen *F. tularensis*, the etiological agent of tularemia, which is vectored by ticks. Advances in next generation sequencing now enable us to assemble the genomes of endosymbionts without first isolating the bacterium in culture, which is a difficult task due to degradation of their genomes (Dale 2003; Pontes & Dale 2006). We assembled the genomes of the FLEs in the hard tick *Amblyomma maculatum* and the soft tick *Ornithodoros moubata* (Gerhart et al. 2016; Gerhart et al. *in preparation*) from whole genome shotgun reads. While this project was in progress, another group described the genome of an FLE in the soft tick *Argas* (*Persicargas*) *arboreus* (previously referred to as *Argas persicus*) Hinrichs et al. 2016). By utilizing the three genomes, we reveal that FLEs evolved from a mammalian pathogen, and are potentially horizontally transferred between ticks. Additionally, our data indicate that the superior metabolic capabilities of FLEs could have resulted in them replacing ancestral CLEs with degraded genomes in several lineages of hard and soft ticks.

## **METHODS**

### **Sequencing and bacterial identification**

A female *O. moubata* procured from the Institute of Parasitology, Czech Republic, and a female *A. maculatum* obtained from Oklahoma State University Tick Rearing Facility were used in this study. DNA was extracted from each tick as follows: Tick outer surface was sterilized using 70% ethanol, and midgut tissue was extracted by dissection followed by gentle pressure applied to the anterior of the tick. Ejected material was washed in PBS, treated with protease, and DNA was extracted using a gDNA kit and protocol (Qiagen). Purified DNA samples were prepared with a standard Illumina adapter library and sequenced using Illumina Hi-Seq 2500 (100 bp, paired-end reads) at Oregon Health and Science University's Massively Parallel Sequencing Shared Resource (MPSSR), yielding approximately 180 million read pairs per sample. Low confidence reads were removed using Trimmomatic (Bolger et al. 2014) with an Illumina clip of four seed mismatches, a 30 bp palindrome clip threshold, and simple clip threshold of 10, using a standard adapter library. Quality trimming was performed using a sliding window of 4 with a score average of 20 and a minimum read length of 50% (50bp). The identity of bacteria present in the tick microbiome were determined after the first round of de novo assembly with IDBA (Peng et al. 2012) utilizing  $\geq 1$  kb contigs identified via Blastn and binned according to the genus of closest hit using MEGAN5's taxonomy browser (Huson et al. 2007).

### **Genome assembly**

Trimmed reads were assembled into contigs using IDBA-UD (Peng et al. 2012), using a

k-mer range of 49:121, a step size of two, and a minimum contig length of 1kb. Putative FLE contigs were identified through Blastn against a library of *Francisella* genome sequences obtained from NCBI. All trimmed reads were mapped to these contigs using bowtie2 and Samtools (Langmead & Salzberg 2012) and then reassembled with IDBA-UD and IDBA-hybrid into final sets of nine and seven contigs for FLE-Om and FLE-Am, respectively. The genomes sequences are publically available on GenBank (accession numbers: LVCE000000000 and LNCT000000000). The completeness of the assembled genomes were examined using a single-copy gene database (Albertsen et al. 2013), yielding identical results for FLE-Om, FLE-Am, *F. persica*, and *F. tularensis*, indicating that both FLE genomes are mostly complete.

### **Genome annotation**

FLE-Om and FLE-Am contigs were annotated using NCBI's prokaryotic genome annotation pipeline, and pseudogenes were verified manually. Protein-coding genes were binned into categories based on their roles in primary metabolism, amino acid and nucleic acid synthesis, or vitamin and cofactor metabolism using KEGG (Manyam et al. 2015) and MEGAN5. Genes present in FLE-Am, FLE-Om and *F. persica* were compared to illustrate similarities and differences in their metabolic capabilities. A list of genes critical to the pathogenicity of *F. tularensis* (Meibom & Charbit 2010; Rowe & Huntley 2015) was used to identify functional and pseudogenized versions of virulence genes present in FLE-Om, FLE-Am and *F. persica*.

### **Pan-genome construction**

The pan-genome of FLEs was constructed using a Blastn database created from the genes

present in all three genomes using a criteria of at least 80% nucleotide identity over a region of at least 90% to be considered homologous. Biochemical, metabolic and cellular functions were assigned to each group by utilizing KEGG pathways viewer in MEGAN. Diagrams and figures were generated using CircOS (Krzywinski et al. 2009) and Adobe Illustrator.

### **Metabolic reconstruction**

Metabolic reconstruction was performed on the groups of genes identified in the pan-genome construction. The groups of shared and unique genes were separately loaded into MEGAN5's KEGG mapping function and the completeness of metabolic pathways was inferred. If a few genes within a larger pathway were missing from the original analysis, it was confirmed or corrected using Blastp searches against the genomes of *F. tularensis tularensis* and *F. tularensis novicida*.

### **Phylogenetic analysis**

In addition to FLE-Om, FLE-Am and *F. persica*, we included 39 fully sequenced *Francisella* genomes to generate a robust phylogenetic tree. Using reciprocal Blastp, a subset of 419 orthologous genes were identified. Sequences were aligned using Clustal Omega (Sievers et al. 2014) and trimmed using Gblocks (Talavera and Castresana 2007). jModelTest2 selected the GTR+I+G (General Time Reversible plus Invariant sites plus Gamma distribution) evolution model (Darriba et al. 2012) as the best fit for the data. Maximum Likelihood and Bayesian trees were constructed using RAxML (Stamatakis 2014) and MrBayes (Huelsenbeck and Ronquist 2001), respectively.

## RESULTS

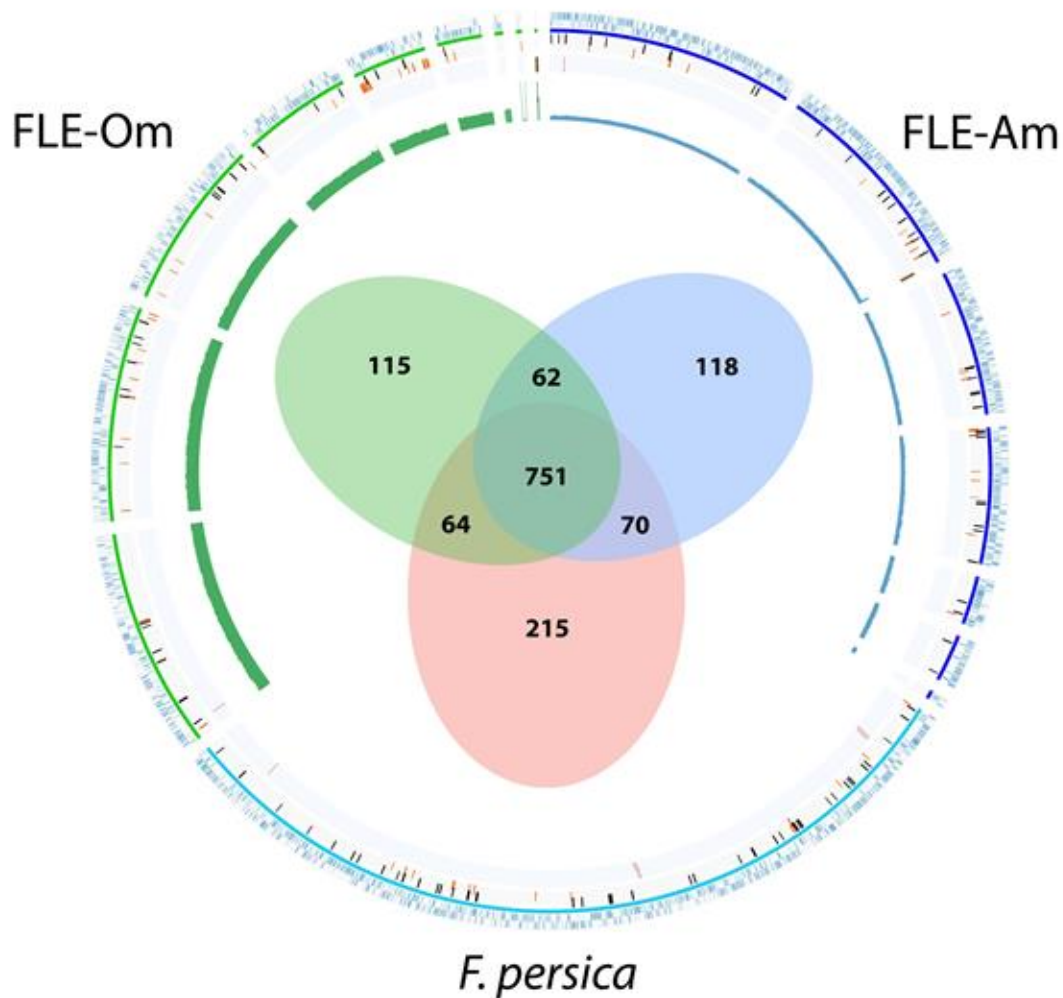
### Genome features of FLEs

The genome of FLE-Am (1.56 Mb) is ~80% the size of the genome of the mammalian pathogen *F. tularensis* (1.89 Mb) (**Table 1**). The nucleotide composition of FLE-Am's genome is very similar to that of pathogenic strains. A significant portion (~33%) of its protein-coding genes contain inactivating mutations (**Figure 1**). Similarly, the genome of FLE-Om is 1.57 Mb, contains similar coding density as FLE-Am (55.5% and 56.9%, respectively) and around 35% of its genes are pseudogenes (**Table 1**). FLE-Om encodes fewer coding genes than FLE-Am, but has a marginally larger genome. The third FLE genome available is of *F. persica* (Hinrichs et al. 2016), and updated by Sjödin *et al.* (NZ\_CP013022.1). Interestingly, unlike FLE-Am and FLE-Om, *F. persica*'s genome (1.54 Mb) contains fewer inactivating mutations, a higher coding density and a longer average gene size (**Table 1**).

**Table 1.** Genome features of *Francisella* species.

<b>Feature</b>	<b>FLE-Om</b>	<b>FLE-Am</b>	<b><i>F. persica</i></b>	<b><i>F. tularensis</i></b>
Contigs	9	7	1	1
Length (bp)	1,574,786	1,556,255	1,540,154	1,892,772
%G+C	31.80%	31.80%	31.39%	32.30%
Coding domains	992	1,001	1,096	1,556
Coding density	55.51%	56.90%	68%	92%
Avg. gene length	881 bp	886 bp	964 bp	937 bp
Pseudogenes	543	484	205	227
Single copy genes <sup>a</sup>	106/111	106/111	106/111	106/111
rRNAs	4	7	9	10
tRNAs	34	32	38	38
Accession #	LVCE000000000	LNCT000000000	NZ_CP013022	NC_006570

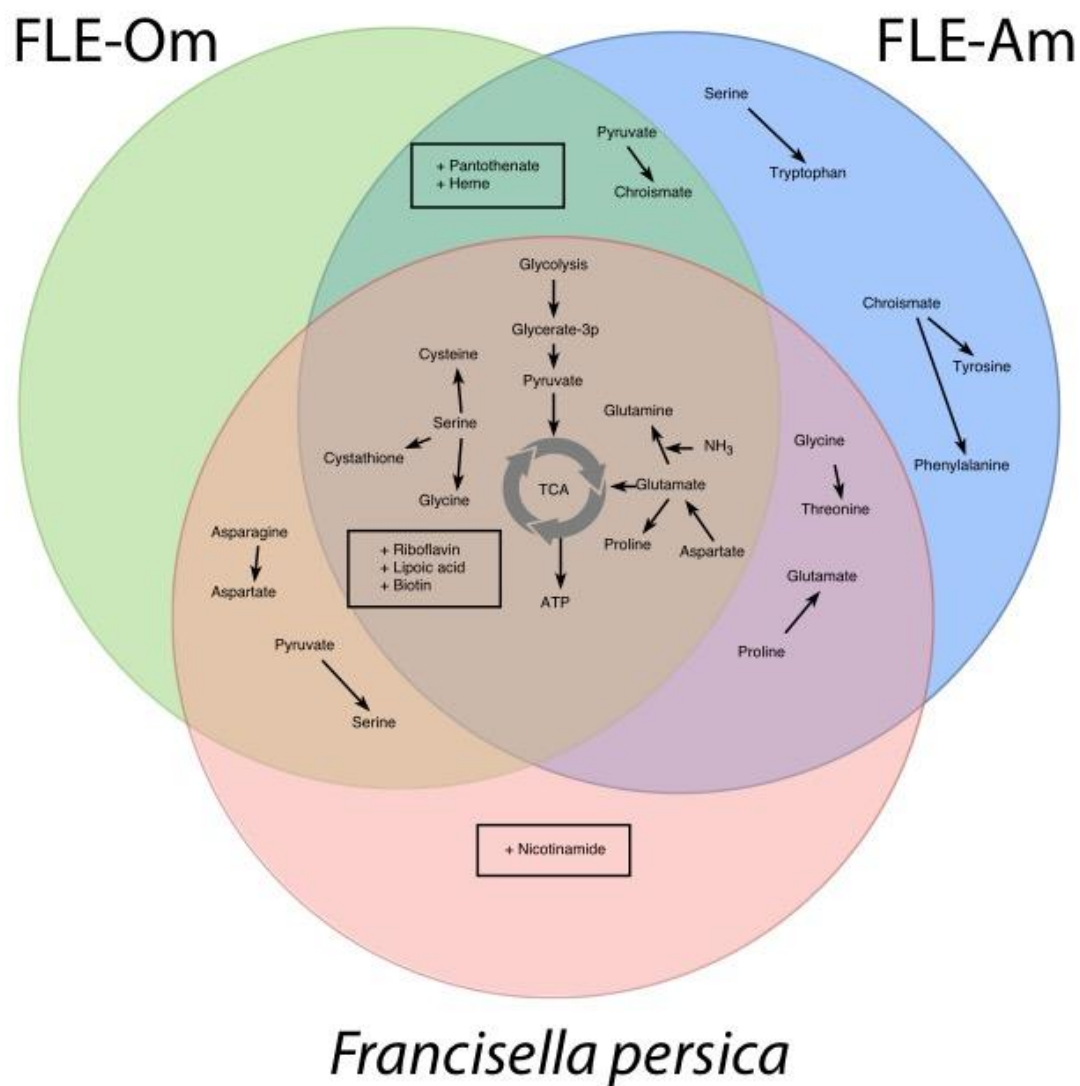
<sup>a</sup> Albertsen et al. 2013



**Figure 1:** Comparison of the three FLE genomes. From the outside: (1) coding domains in both forward and reverse directions in blue, (2) pathogenic genes in orange and pseudogenized pathogenic genes in black, (3) identified ribosomal RNA genes, and (4) read coverage for FLE-Om and FLE-Am in green and blue, respectively. The inner Venn diagram depicts the FLE pan-genome, showing conserved, peripheral and unique genes in each bacterium.



The three FLE genomes were used to determine their pan-genome, which consists of 1395 genes, with the majority (751 genes) in the core genome common between all three strains. In addition, all three genomes show clear evidence of their pathogenic heritage, encoding functional and pseudogenized virulence genes (**Figure 1**). We examined the metabolic potential of the FLEs (**Figure 2**) and found that *F. persica* encodes a large, unique biosynthetic pathway for nicotinamide, while lacking complete pathways for pantothenate and heme that are present in FLE-Om and FLE-Am. FLE-Am produces the most diverse group of amino acids, while FLE-Om has no unique pathways. All endosymbionts have a TCA cycle & electron transport chain, a glutamine/glutamate synthesis pathway that is a potential waste nitrogen processing pathway, and cysteine, cystathione and glycine conversion enzymes. Within the auxiliary genomes, serine and chorismate synthesis pathways are shared between FLE-Om and FLE-Am, whereas glutamate semialdehyde dehydrogenase and asparagine deaminase are shared between FLE-Am and *F. persica*. Furthermore, FLE-Om shares a larger number of pseudogenes with FLE-Am than with *Francisella persica*. Collectively, our data show that while all three FLEs have the potential to provide several critical nutrients to their respective hosts, FLE-Om and FLE-Am seem to encode similar functionalities.

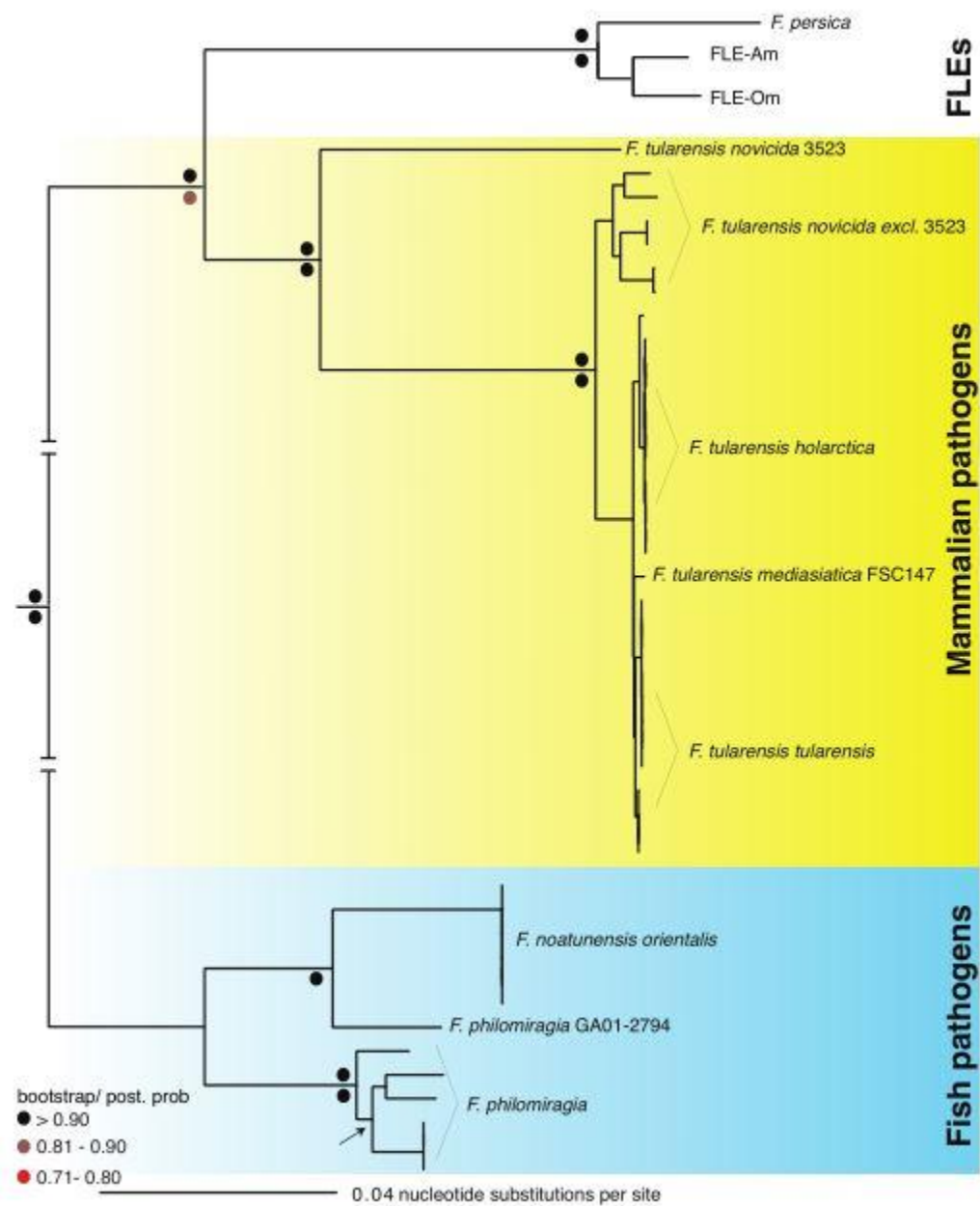


**Figure 2:** Comparison of metabolic features of FLEs. Arrows indicate simple one- or two-enzyme processes, boxes indicate larger, multi-step biosynthetic pathways.

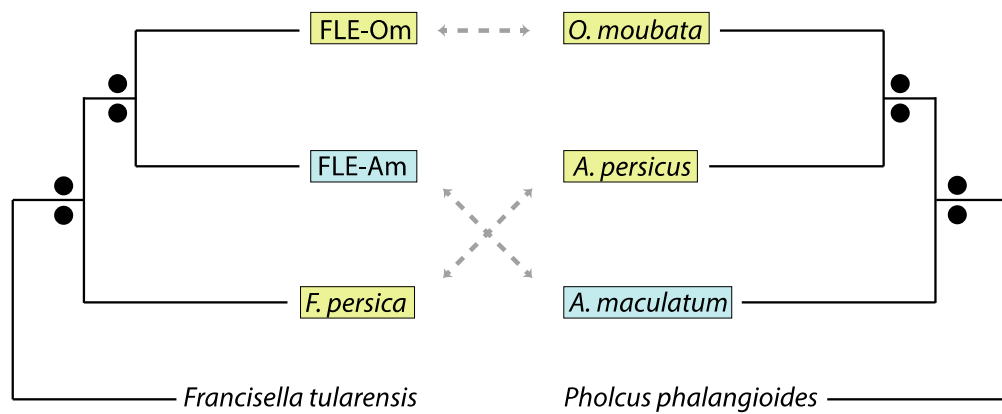
### **FLEs descended from a common pathogenic ancestor**

I generated a *Francisella* phylogeny using the complete orthologous gene content of all available *Francisella* genomes, including FLE-Am, FLE-Om and *F. persica*. The final gene set contains 419 single copy orthologous genes, and shows that all FLEs evolved from a common ancestor related to the mammalian pathogen *F. tularensis* (**Figure 3**). Further, the data also suggest that horizontal transfer of endosymbionts likely occur in ticks: the FLE in the soft tick *O. moubata* is more closely related to the FLE in *A. maculatum*, a hard tick, than to the FLE in the other soft tick *Argas (Persicargas) arboreus* (also referred to as *Argas persicus*) (**Figure 4**).

To confirm the relationship between pathogenic *Francisella* and FLEs, I examined the FLE genomes for the presence of virulence genes described in *F. tularensis* (Thomas et al. 2007; Meibom & Charbit 2010; Rowe & Huntley 2015). Both FLE-Am and FLE-Om contained pseudogenized versions of several virulence genes, including a Type VI Secretion System present on a pathogenicity island in *F. tularensis*, and in a Type 4 pili that is critical to mammalian infection (**Table 2**). In contrast, the *F. persica* genome contained a higher number of potentially functional virulence genes, including important systems such as lipopolysaccharide (LPS) and the structural and anchoring components of the Type VI secretion system, suggesting that it is still pathogenic to mammals. In fact, *F. persica* (previously wrongly identified as *Wolbachia persica*) was found to cause fever and death in guinea pigs (Suitor & Weiss 1961).



**Figure 3:** Phylogeny of FLEs using 419 single copy orthologous genes conserved in 39 fully sequenced strains of *Francisella*. FLEs appear as a sister taxon to *Francisella tularensis*, a clade pathogenic to mammals.



**Figure 4:** Comparison of FLE and host phylogenies. Phylogenetic placement of FLEs based on 419 orthologous genes (left) do not correspond to the host phylogeny (right) based on an 18S rDNA sequences, suggesting that FLEs in soft ticks did not diverge along with their hosts, but instead were acquired recently from another host or from the environment. Soft ticks and their FLEs are highlighted in yellow boxes. The hard tick and its FLE are in blue boxes. Black dots indicate >90% bootstrap value (top) or posterior probability (bottom).

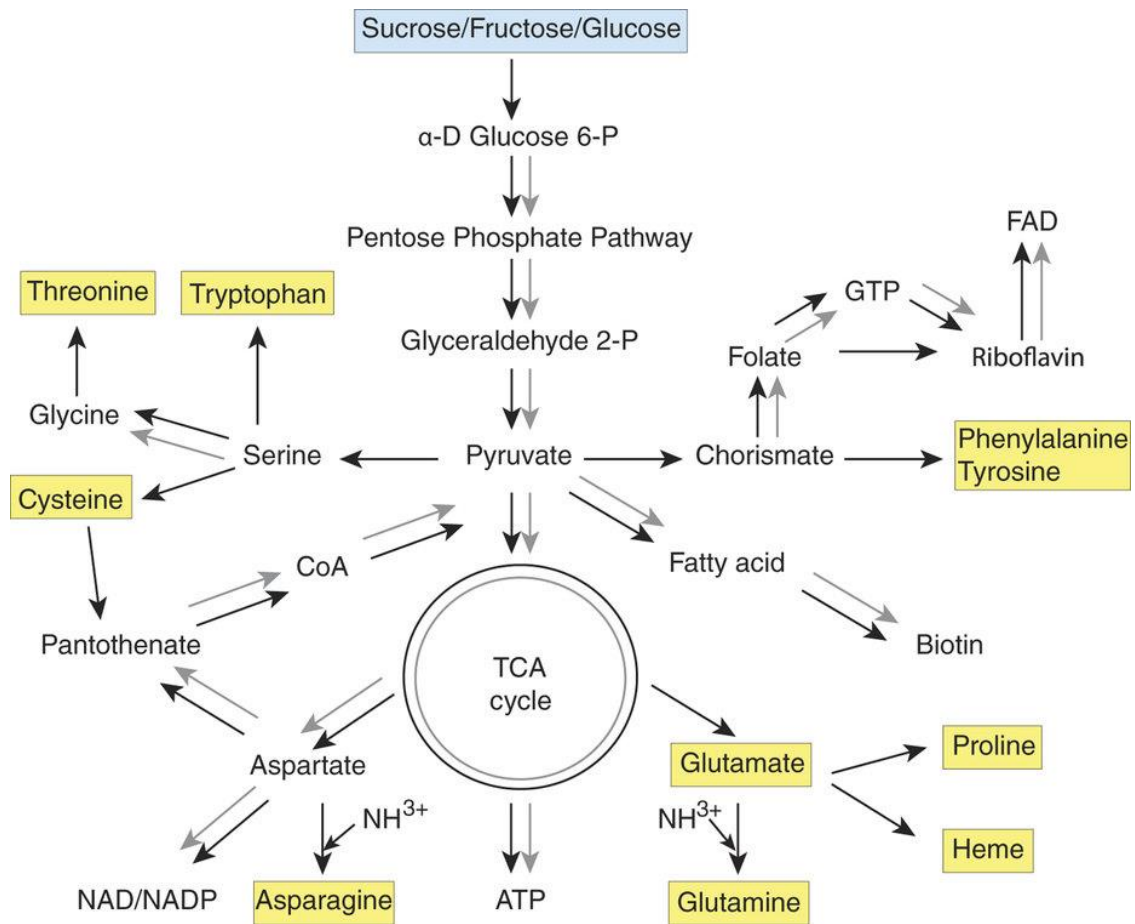
**Table 2.** Virulence genes in FLEs compared to *F. tularensis*.

	<i>F. tularensis</i>		<i>F. persica</i>		FLE-Am		FLE-Om	
	Coding genes	Pseudo genes	Coding genes	Pseudo genes	Coding genes	Pseudo genes	Coding genes	Pseudo genes
LPS, capsule & membrane synthesis	22	0	13	1	8	6	9	2
Antigen synthesis	11	1	5	2	6	1	5	3
Type 4 Pili	17	0	8	4	4	10	2	6
Hypothetical porins	2	0	2	0	1	3	1	0
Outer membrane proteins	8	0	8	0	6	2	5	0
Inner membrane proteins	10	0	8	0	8	0	8	1
Type VI Secretion System	17	0	16	0	3	4	1	6

## DISCUSSION

### FLEs could potentially replace ancestral CLEs

The geographical range of *A. maculatum* (Gulf coast tick) overlaps with that of *A. americanum* (Lone Star tick) in the southwest United States. Interestingly, the primary endosymbiont of *A. americanum* is a CLE, termed CLEAA (Smith et al. 2015) and not an FLE, as observed in *A. maculatum* (FLE-Am). Bacteria with extensive metabolic proficiency are known to replace ancient endosymbionts with reduced metabolic prowess (Koga & Moran 2014; Bennette and Moran 2015), hence we compared the metabolic pathways present in CLEAA and FLE-Am. As shown in **Figure 5**, the metabolic capability of FLE-Am is much more extensive than in CLEAA. For instance, FLE-Am can produce heme in addition to every cofactor synthesized by CLEAA, and while both FLE-Am and CLEAA share the ability to produce aspartate from pyruvate and to metabolize it into ATP, only the FLE synthesizes cysteine, threonine, tyrosine, tryptophan, and phenylalanine, and can synthesize serine from pyruvate, and metabolizes glutamate, glutamine, and asparagine into ATP. Of these amino acids, FLE-Am most crucially provides a reliable source of the amino acid cysteine, which is found in very low concentrations in bovine blood (Duarte et al. 1999). FLE-Am can also synthesize glutamine from glutamic acid and ammonia, thus recycling cellular ammonia waste to useful products (Sabree et al. 2009). In sum, our data suggest that the superior biosynthetic capability of FLE-Am confers a selective advantage, which could have led to FLE-Am recently replacing an ancestral symbiont such as CLE with reduced metabolic capacity in *A. maculatum*.



**Figure 5:** Comparison of metabolic potentials of FLE-Am (black arrows) and CLEAA (grey arrows). FLE-Am is capable of synthesizing every major metabolite that CLEAA can provide to the host, along with several other potentially beneficial compounds that CLEAA is unable to provide (yellow boxes).



This scenario is supported by the observation that CLEs are much more prevalent in ticks than FLEs but appear to have been competitively excluded by FLEs in several tick species (Duron et al. 2017).

Our phylogenetic analyses indicate that FLEs share a recent common ancestor with *Francisella* species that are pathogenic to mammals, suggesting that FLEs likely arose by the domestication of a mammalian pathogen that was vectored by ticks. This process could occur rapidly because the bacterium does not have to ‘learn’ anew how to circumvent the tick’s immune response and there is no need to attenuate bacterial virulence towards the tick (Clayton et al. 2012). Additionally, because genes that promote pathogenesis through amino acid scavenging (e.g., FTT 0968c, *xasA*), macrophage survival (e.g., *carA*, *carB*, *bioF*), and intracellular replication (*aroEI*, *purMCD*, *purL*, and *purF*) (Rowe & Huntley 2015; Meibom & Charbit 2010) are retained in FLEs, they could be key to its endosymbiotic lifestyle. Contrasting this, in *Coxiella*, previous studies indicate that pathogen-symbiont transition occurred in the reverse direction: the mammalian pathogen *C. burnetii* probably originated from a tick-associated ancestor, which acquired genes required for successfully infecting mammalian cells via horizontal gene transfer (Duron et al. 2015; Moses et al. 2017).

Previous studies have shown that the provisioning of nutrients such as amino acids and vitamins, or processing waste nitrogen are some of the major roles of endosymbionts (Sabree et al. 2009; Hansen & Moran 2011; Nikoh et al. 2014; Wilson & Duncan 2015). We examined the metabolic pathways present in FLEs and found that they

have the ability to both synthesize several amino acids and coenzymes present in very low concentrations in blood, and to recycle ammonia into amino acids (**Figure 2, Figure 5**), indicating that they could improve host fitness by functioning as nutrient provisioning endosymbionts (Gerhart et al. 2016; Gerhart et al. *in preparation*). The extensive metabolic capabilities of FLEs could be a selective advantage over other endosymbionts with limited metabolic potential such as ancestral CLEs with degraded genomes, thereby resulting in FLEs replacing older endosymbionts (Gerhart et al. 2016; Duron et al. 2017).

Unlike congruent host and symbiont phylogenetic trees observed in many insects (e.g. Baumann et al. 1997; Russell & Moran 2005), host-endosymbiont coevolution is rare across ticks, probably due to horizontal transfer of symbionts between unrelated tick species. A recent survey of 81 species of hard and soft ticks showed that FLEs in several soft ticks were more closely related to FLEs in hard ticks than to each other, suggesting that FLEs were acquired multiple times by ticks (Duron et al. 2017). In accordance with this observation, it is likely that the two FLEs in soft ticks analyzed in our study were acquired independently: FLE-Om from a hard tick, and *F. persica* from an unknown source. Horizontal transfer of endosymbionts appears to be more prevalent in ticks than in insects. This is probably because tick endosymbionts are not restricted to specialized cells as seen in insects, but are present in several tissues including in the salivary glands (Budachetri et al. 2014), which could facilitate transfer between blood-feeding arthropods while co-feeding on the same vertebrate host (Wright et al. 2015; Rudolf et al. 2016).

In conclusion, we describe both the unusual genomic features of three endosymbionts of genus *Francisella*, and provide many avenues for future research in the

field of tick endosymbiosis, which is an understudied area with important implications to human health. Going forward, more extensive sampling and careful sequencing will be required to deduce the ancestor of FLEs, to describe FLE diversity across ticks, to confirm horizontal transfer of FLEs, and to test the competitive model of endosymbiont replacement in ticks.

## REFERENCES

- Ahantarig A et al. 2013. Hard ticks and their bacterial endosymbionts (or would be pathogens). *Folia Microbiol.* 58:419–428. doi: 10.1007/s12223-013-0222-1.
- Albertsen M et al. 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol.* 31:533–538. doi: 10.1038/nbt.2579.
- Baumann P. 2005. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annu Rev Microbiol.* 59:155–189. doi: 10.1146/annurev.micro.59.030804.121041.
- Baumann P, Moran NA, Baumann L. 1997. The evolution and genetics of aphid endosymbionts. *Bioscience* 47:12–20. doi: 10.2307/1313002.
- Bennett GM, Moran NA. 2015. Heritable symbiosis: the advantages and perils of an evolutionary rabbit hole. *Proc Natl Acad Sci USA.* 112:10169–10176. doi: 10.1073/pnas.1421388112.
- Black WC, Piesman J. 1994. Phylogeny of hard- and soft-tick taxa (Acari: Ixodida) based on mitochondrial 16S rDNA sequences. *Proc Natl Acad Sci USA.* 91:10034–10038. doi: 10.1073/pnas.91.21.10034.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. doi: 10.1093/bioinformatics/btu170.
- Bright M, Bulgheresi S. 2010. A complex journey: transmission of microbial symbionts. *Nat Rev Microbiol.* 8:218–230. doi: 10.1038/nrmicro2262.

- Budachetri K et al. 2014. An insight into the microbiome of the *Amblyomma maculatum* (Acari: Ixodidae). *J Med Entomol.* 51:119–29. doi: 10.1603/ME12223.
- Burke GR, Moran NA. 2011. Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biol Evol.* 3:195–208. doi: 10.1093/gbe/evr002.
- Challacombe JF et al. 2017. Whole-genome relationships among *Francisella* bacteria of diverse origins define new species and provide specific regions for detection. *Appl Environ Microbiol.* 83(3):e02589-16 (2017). doi: 10.1128/AEM.00174-17
- Clayton AL et al. 2012. A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect–bacterial symbioses. *PLoS Genet.* 8:e1002990. doi: 10.1371/journal.pgen.1002990.
- Dale C. 2003. Loss of DNA recombinational repair enzymes in the initial stages of genome degeneration. *Mol Biol Evol.* 20:1188–1194. doi: 10.1093/molbev/msg138.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772–772. doi: 10.1038/nmeth.2109.
- Duarte RT, Carvalho Simões MC, Sgarbieri VC. 1999. Bovine blood components: fractionation, composition, and nutritive value. *J Agric Food Chem.* 47:231–236. doi: 10.1021/jf9806255.

- Duron O, Jourdain E, McCoy KD. 2014. Diversity and global distribution of the *Coxiella* intracellular bacterium in seabird ticks. *Ticks Tick-borne Dis.* 5:557–563. doi: 10.1016/j.ttbdis.2014.04.003.
- Duron O et al. 2015. The recent evolution of a maternally-inherited endosymbiont of ticks led to the emergence of the Q fever pathogen, *Coxiella burnetii*. *PLoS Pathog.* 11:e1004892. doi: 10.1371/journal.ppat.1004892.
- Duron O et al. 2017. Evolutionary changes in symbiont community structure in ticks. *Mol Ecol.* 38:42–49. doi: 10.1111/mec.14094.
- Fitzpatrick J. 2013. Global food security: the impact of veterinary parasites and parasitologists. *Vet Parasitol.* 195:233–248. doi: 10.1016/j.vetpar.2013.04.005.
- Gerhart JG, Moses AS, Raghavan R. 2016. A *Francisella*-like endosymbiont in the Gulf Coast tick evolved from a mammalian pathogen. *Sci Rep.* 6:33670. doi: 10.1038/srep33670.
- Hansen AK, Moran NA. 2011. Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proc Natl Acad Sci USA.* 108:2849–2854. doi: 10.1073/pnas.1013465108.
- Hinrichs SH et al. 2016. Reclassification of *Wolbachia persica* as *Francisella persica* comb. nov. and emended description of the family Francisellaceae. *Int J Syst Evol Microbiol.* 66:1200–1205. doi: 10.1099/ijsem.0.000855.
- Hooper LV. 2001. Commensal host-bacterial relationships in the gut. *Science* 292:1115–1118. doi: 10.1126/science.1058709.

- Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755. doi: 10.1093/bioinformatics/17.8.754.
- Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. *Genome Res.* 17:377–386. doi: 10.1101/gr.5969107.
- Jongejan F, Uilenberg G. 2005. The global importance of ticks. *Parasitology* 129:S3. doi: 10.1017/S0031182004005967.
- Klompen J. 2000. Systematics and biogeography of hard ticks, a total evidence approach. *Cladistics* 16:70–102. doi: 10.1006/clad.1999.0126.
- Klyachko O et al. 2007. Localization and visualization of a *Coxiella*-type symbiont within the lone star tick, *Amblyomma americanum*. *Appl Environ Microbiol.* 73:6584–6594. doi: 10.1128/AEM.00537-07.
- Koga R, Moran NA. 2014. Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont. *ISME J.* 8:1237–1246. doi: 10.1038/ismej.2013.235.
- Krzywinski M et al. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res.* 19:1639–1645. doi: 10.1101/gr.092759.109.
- Lamelas A et al. 2011. *Serratia symbiotica* from the aphid *Cinara cedri*: a missing link from facultative to obligate insect endosymbiont. *PLoS Genet.* 7:e1002357. doi: 10.1371/journal.pgen.1002357.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. doi: 10.1038/nmeth.1923.
- Manyam G, Birerdinc A, Baranova A. 2015. KPP: KEGG pathway painter. *BMC*

- Syst Biol.* 9:S3. doi: 10.1186/1752-0509-9-S2-S3.
- Meibom KL, Charbit A. 2010. The unraveling panoply of *Francisella tularensis* virulence attributes. *Curr Opin Microbiol.* 13:11–17. doi: 10.1016/j.mib.2009.11.007.
- Moran NA. 1996. Accelerated evolution and Muller’s ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci USA.* 93:2873–2878. doi: 10.1073/pnas.93.7.2873.
- Moreira D, López-García P. 1998. Symbiosis between methanogenic archaea and  $\delta$ -proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J Mol Evol.* 47:517–530. doi: 10.1007/PL00006408.
- Nava S, Guglielmone AA, Mangold AJ. 2009. An overview of systematics and evolution of ticks. *Front Biosci (Landmark Ed.)* 14:2857–77. doi: 10.2735/3418.
- Nikoh N et al. 2014. Evolutionary origin of insect-*Wolbachia* nutritional mutualism. *Proc Natl Acad Sci USA.* 111:10257–10262. doi: 10.1073/pnas.1409284111.
- Noda H, Munderloh UG, Kurtti TJ. 1997. Endosymbionts of ticks and their relationship to *Wolbachia* spp. and tick-borne pathogens of humans and animals. *Appl Environ Microbiol.* 63:3926–3932. doi: 0099-2240/97/\$04.00+0.
- Nyholm S V, McFall-Ngai M. 2004. The winnowing: establishing the squid-*Vibrio* symbiosis. *Nat Rev Microbiol.* 2:632–642. doi: 10.1038/nrmicro957.
- Peng Y, Leung HCM, Yiu SM, Chin FYL. 2012. IDBA-UD: a de novo assembler for

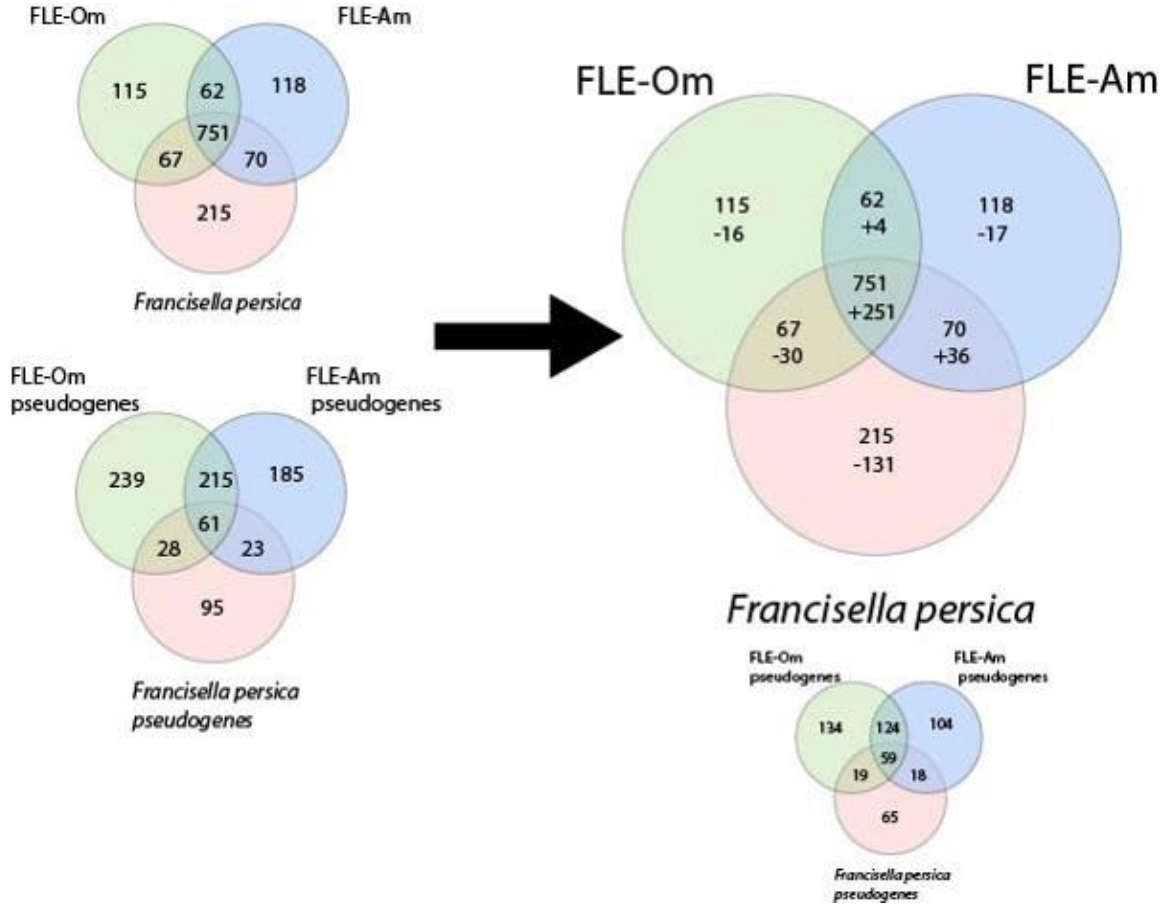


- single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28:1420–1428. doi: 10.1093/bioinformatics/bts174.
- Pontes MH, Dale C. 2006. Culture and manipulation of insect facultative symbionts. *Trends Microbiol.* 14:406–412. doi: 10.1016/j.tim.2006.07.004.
- Randolph SE, Gern L, Nuttall PA. 1996. Co-feeding ticks: epidemiological significance for tick-borne pathogen transmission. *Parasitol Today* 12:472–479. doi: 10.1016/S0169-4758(96)10072-7.
- Rowe HM, Huntley JF. 2015. From the outside-in: the *Francisella tularensis* envelope and virulence. *Front Cell Infect Microbiol.* 5:1–20. doi: 10.3389/fcimb.2015.00094.
- Rudolf I et al. 2016. Molecular survey of arthropod-borne pathogens in sheep keds (*Melophagus ovinus*), central europe. *Parasitol Res.* 115:3679–3682. doi: 10.1007/s00436-016-5175-2.
- Russell J, Moran N. 2005. Horizontal transfer of bacterial symbionts: heritability and fitness effects in a novel aphid host. *Appl Environ Microbiol.* 71:7987–7994. doi: 10.1128/AEM.71.12.7987-7994.2005.
- Sabree ZL, Kambhampati S, Moran NA. 2009. Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endosymbiont. *Proc Natl Acad Sci USA.* 106:19521–19526. doi: 10.1073/pnas.0907504106.
- Scoles GA. 2004. Phylogenetic analysis of the *Francisella*-like endosymbionts of *Dermacentor* ticks. *J Med Entomol.* 41:277–286. doi: 10.1603/0022-2585-41.3.277.

- Sievers F et al. 2014. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol Syst Biol.* 7:539–539. doi: 10.1038/msb.2011.75.
- Smith TA, Driscoll T, Gillespie JJ, Raghavan R. 2015. A *Coxiella*-like endosymbiont is a potential vitamin source for the Lone Star tick. *Genome Biol Evol.* 7:831–838. doi: 10.1093/gbe/evv016.
- Spielman A, Ribeiro JMC, Mather TNN, Piesman J. 1987. Dissemination and salivary delivery of lyme disease spirochetes in vector ticks (Acari: Ixodidae). *J Med Entomol.* 24:201–205. doi: 10.1093/jmedent/24.2.201.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. doi: 10.1093/bioinformatics/btu033.
- Suitor EC, Weiss E. 1961. Isolation of a *Rickettsia*-like microorganism (*Wolbachia persica*, n. sp.) from *Argas persicus* (Oken). *J Infect Dis.* 108:95–106. doi: 10.1093/infdis/108.1.95.
- Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 56:564–577. doi: 10.1080/10635150701472164.
- Thomas RM et al. 2007. The immunologically distinct O antigens from *Francisella tularensis* subspecies *tularensis* and *Francisella novicida* are both virulence determinants and protective antigens. *Infect Immun.* 75:371–378. doi: 10.1128/IAI.01241-06.

- Weiss BL, Wang J, Aksoy S. 2011. Tsetse immune system maturation requires the presence of obligate symbionts in larvae. *PLoS Biol.* 9:e1000619. doi: 10.1371/journal.pbio.1000619.
- Wilson ACC, Duncan RP. 2015. Signatures of host/symbiont genome coevolution in insect nutritional endosymbioses. *Proc Natl Acad Sci USA.* 112:10255–10261. doi: 10.1073/pnas.1423305112.
- Wright CL, Sonenshine DE, Gaff HD, Hynes WL. 2015. *Rickettsia parkeri* transmission to *Amblyomma americanum* by cofeeding with *Amblyomma maculatum* (Acari: Ixodidae) and potential for spillover. *J Med Entomol.* 52:1090–1095. doi: 10.1093/jme/tjv086.

## APPENDIX A



Pseudogene analysis. Conserved pseudogenes between the three FLEs are represented in the bottom left. These conserved pseudogenes can be compared to the pan-genome to better understand the differences between the three FLE strains (right). This ‘pseudogene-pan-genome’ shows that FLEs have undergone differential pseudogenization, but were once very similar.

# SCIENTIFIC REPORTS

OPEN

## A *Francisella*-like endosymbiont in the Gulf Coast tick evolved from a mammalian pathogen

Received: 27 June 2016

Accepted: 30 August 2016

Published: 20 September 2016

Jonathan G. Gerhart, Abraham S. Moses &amp; Rahul Raghavan

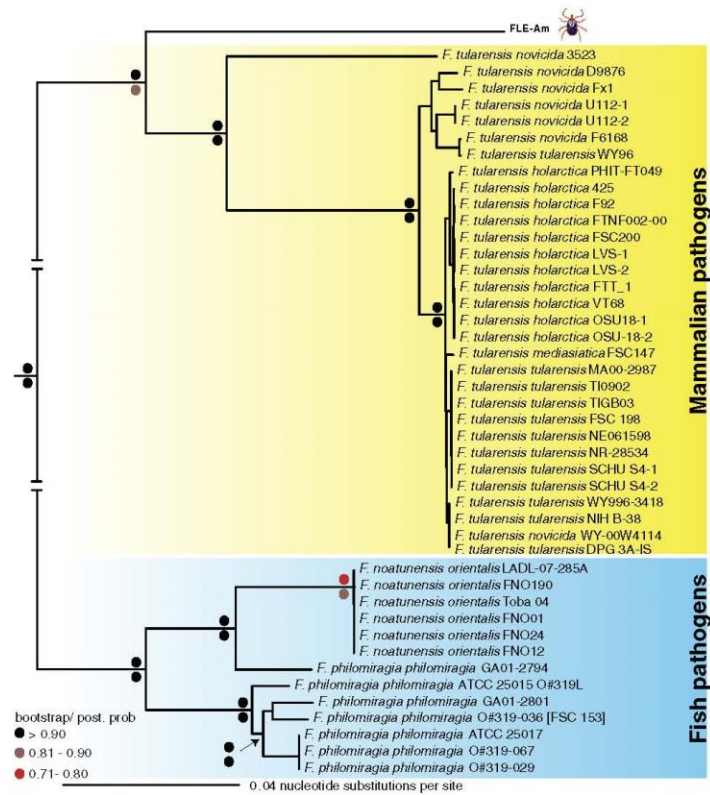
Ticks (order Ixodida) vector pathogenic bacteria that cause diseases in humans and other mammals. They also contain bacteria that are closely related to pathogens but function as endosymbionts that provide nutrients that are missing from mammalian blood—their sole food source. For instance, mammalian pathogens such as *Coxiella burnetii* and *Francisella tularensis*, as well as *Coxiella*-like and *Francisella*-like endosymbionts (CLEs and FLEs, respectively) occur in ticks worldwide. However, it is not clear whether the pathogens evolved from symbionts or symbionts from pathogens. Recent studies have indicated that *C. burnetii* likely originated from a tick-associated ancestor, but the origins of FLEs are not clear. In this study, we sequenced the genome of an FLE, termed FLE-Am, present in the Gulf Coast tick, *Amblyomma maculatum*. We show that FLE-Am likely evolved from a pathogenic strain of *Francisella*, indicating that tick endosymbionts can evolve from mammalian pathogens. Although the genome of FLE-Am is almost the same size as the genomes of pathogenic *Francisella* strains, about one-third of its protein-coding genes contain inactivating mutations. The relatively low coding capacity and extensive metabolic capabilities indicate that FLE-Am transitioned recently to its current endosymbiotic lifestyle and likely replaced an ancient endosymbiont with degraded functionality.

Blood is the sole source of nutrition for ticks; however, it contains minimal amounts of several amino acids and cofactors required for normal development<sup>1</sup>. Similar to insects that thrive on unbalanced plant diets<sup>2</sup>, ticks depend on symbiotic bacteria such as *Coxiella*-like and *Francisella*-like endosymbionts (CLE and FLE, respectively) that likely provision nutrients missing from their diet<sup>3,4</sup>. In addition to containing endosymbionts, ticks often vector closely related mammalian pathogens, including *Francisella tularensis* and *Coxiella burnetii*<sup>5,6</sup>, but the evolutionary relationship between the two clades is not clear. Among *Amblyomma* species (hard-backed tick, family Ixodidae), the most prevalent endosymbionts are CLEs, which have reduced genomes, a feature associated with long-term endosymbiosis<sup>7</sup>. Interestingly, unlike most other *Amblyomma* species, the primary endosymbiont of *A. maculatum* (Gulf Coast tick) is a FLE, termed FLE-Am<sup>8</sup>. We sequenced its genome and show that it likely evolved from a pathogenic ancestor, thereby showing that tick endosymbionts could evolve from mammalian pathogens. Unlike the highly reduced genomes of CLEs, FLE-Am has undergone minimal genome reduction, but a substantial number of its protein-coding genes, including virulence genes, contain inactivating mutations. This low coding capacity indicates that the endosymbiont originated recently, and additionally, by virtue of its superior metabolic and biosynthetic capabilities, FLE-Am most likely replaced an ancestral CLE with degraded functionality, thereby allowing *A. maculatum* to escape the “symbiosis rabbit hole”<sup>9</sup>.

### Results and Discussion

**FLE-AM evolved from a pathogenic ancestor.** We analyzed the *A. maculatum* microbiome and discovered that FLE-Am dominates the tick microbiome. Out of the 4,603 (>1 kb) contigs assembled from the sequencing reads, a vast majority was of host origin. Only 48 contigs were of bacterial origin, and out of those 46 belonged to Francisellaceae (Supplementary Table S1). FLE-Am has been previously shown to be the major bacterium present in *A. maculatum*<sup>8</sup>, and FLEs are maternally transmitted<sup>10</sup>, indicating that it is essential to the normal development of *A. maculatum*. We estimated the phylogenetic position of FLE-Am using 442 orthologous proteins present in 44 fully sequenced *Francisella* genomes. As shown in Fig. 1, FLE-Am shares a recent common ancestor with *Francisella* species that are pathogenic to mammals, with the exclusion of aquatic *Francisella* species

Department of Biology and Center for Life in Extreme Environments, Portland State University, Portland, Oregon, 97201, USA. Correspondence and requests for materials should be addressed to R.R. (email: rahul.raghavan@pdx.edu)



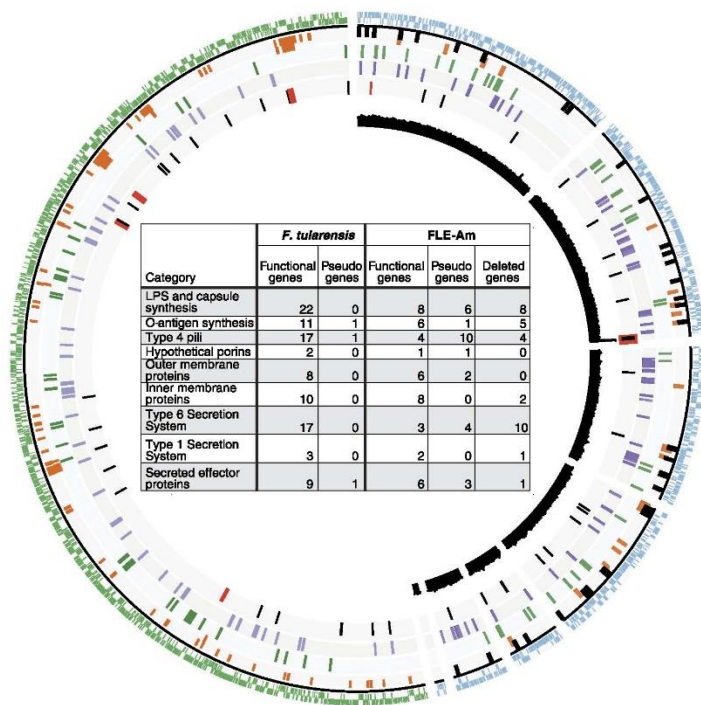
**Figure 1.** FLE-Am is a sister taxon of mammalian pathogens. Phylogenetic tree based on 442 orthologous genes in FLE-Am and 44 fully sequenced *Francisella* genomes is shown. Bootstrap and posterior probability values are provided on top and bottom of each node, respectively.

that are pathogenic to fish. Our multi-protein tree clarifies earlier phylogenetic trees based on partial FLE-Am 16S rRNA sequences that showed it to be a sister taxon of *F. tularensis*<sup>11</sup>.

**Virulence genes have been inactivated in FLE-AM.** As shown recently for *C. burnetii*, mammalian pathogens could evolve from non-pathogenic ancestors by acquiring virulence genes<sup>12</sup>; conversely, avirulent symbionts could arise from pathogenic ancestors by losing virulence genes, but no clear examples of this process have been documented. To identify the evolutionary relationship between pathogenic *Francisella* and FLE-Am, we examined the FLE-Am genome for the presence of virulence genes described in *F. tularensis* and *F. novicida*<sup>13,14</sup>. We discovered that FLE-Am contained pseudogenized versions of several virulence genes, including genes for a Type VI Secretion System present on a pathogenicity island in *F. tularensis* and for Type 4 pili that are critical to mammalian infection (Fig. 2, Supplementary Table S2). Collectively, our data denotes that the ancestor of FLE-Am was most likely a mammalian pathogen that contained functional versions of virulence genes. The absence of intact secretion and effector gene systems in FLE-Am suggests that it is avirulent to humans despite its presence in the salivary glands and saliva of *A. maculatum*<sup>9</sup>. Intriguingly, salivary glands of other *Amblyomma* species such as *A. americanum* contain CLEs<sup>15</sup>. Although its functional significance is not understood, being secreted in saliva could facilitate the exchange of endosymbionts between ticks while co-feeding on the same host<sup>16</sup>.

**FLE-Am evolved recently and likely replaced an ancestral endosymbiont.** The genome of FLE-Am (1.56 Mb) is ~80% the size of the genome of the mammalian pathogen *F. tularensis* (1.89 Mb) (Table 1), and a significant portion (~33%) of its protein-coding genes contains inactivating mutations (Fig. 2, Supplementary Dataset 1). The degree of reductive genome evolution in FLE-Am is much lower than what is usually observed



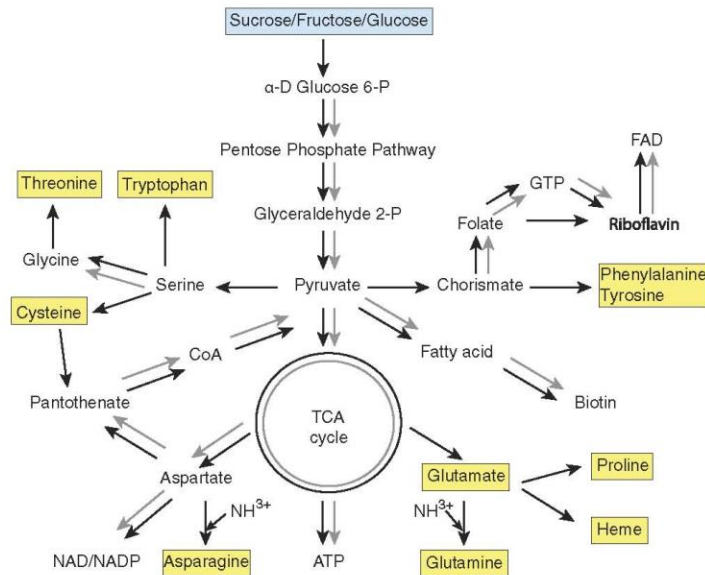


**Figure 2. Comparison of FLE-Am and *F. tularensis* genomes.** FLE-Am genome is on the right and *F. tularensis* SCHU-S4 genome is on the left. The outer two rings show protein-coding genes (blue in FLE-Am and green in *F. tularensis*). Rings 3 and 4 contain virulence genes (intact genes in orange and pseudogenes in black). Amino acid biosynthesis genes (green) are in ring 5. Cofactor synthesis and transport genes (purple) are in ring 6, and tRNAs (black) and rRNAs (red) are in ring 7. Sequencing read coverage for FLE-Am is shown in the innermost black semi-circle. Notice that the coverage for the rRNA operon is double that of the rest of the genome. The table in the center highlights the inactivation and loss of virulence genes in FLE-Am.

	<i>F. tularensis</i>	FLE-Am
Genome size (bp)	1,892,772	1,556,255
GC%	32.3	31.8
Protein coding genes	1,556	1,001
Pseudogenes	227	484
rRNAs	10	7
tRNAs	38	32
Single copy genes	106/111	106/111
Average Gene Size	942 bp	886 bp
GenBank Accession	NC_006570.2	LNCT00000000

**Table 1. Genome features of FLE-Am and *F. tularensis* SCHU-S4.**

in long-term endosymbionts such as a CLE in *A. americanum* (CLEAA)<sup>3</sup>, and the presence of large number of pseudogenes imply that the bacterium is in the initial stages of reductive evolution, as superfluous genes are first pseudogenized and then ultimately deleted from the genome when a bacterium shifts from a free-living to a host-associated lifestyle. Additionally, endosymbiont genomes tend to be much more A+T biased than closely related pathogens or environmental bacteria; however, the nucleotide composition (G+C%) of FLE-Am is very



**Figure 3.** Comparison of metabolic capabilities of FLE-Am and CLEAA. Black arrows mark pathways present in FLE-Am, and grey arrows depict CLEAA pathways. Metabolic products exclusively produced by FLE-Am are highlighted in yellow.

similar to that of *F. tularensis* (32.3% and 31.8%, respectively)<sup>7,9</sup>. These features collectively indicate that FLE-Am transformed recently into an endosymbiont.

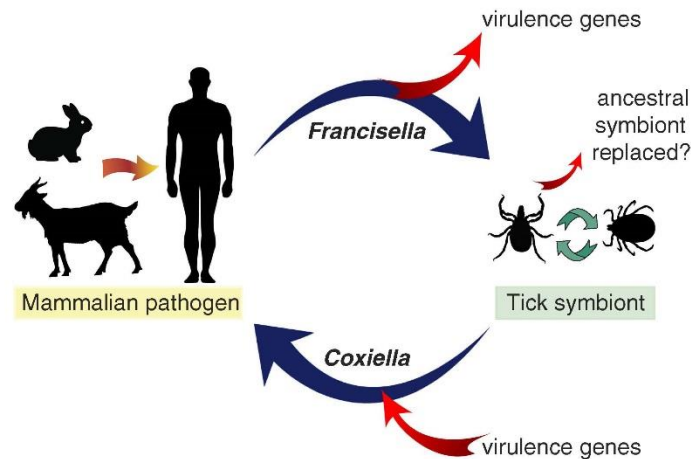
Because bacteria with extensive metabolic proficiency are known to replace ancient endosymbionts with reduced metabolic prowess<sup>2</sup>, we compared the metabolic pathways present in FLE-Am to that of CLEAA, the cofactor provisioning endosymbiont found in *A. americanum*<sup>5</sup>. As shown in Fig. 3, the metabolic capability of FLE-Am is much more extensive than in CLEAA. For instance, FLE-Am can produce heme in addition to cofactors (except thiamine) synthesized by CLEAA. Furthermore, while both FLE-Am and CLEAA share the ability to produce aspartate from pyruvate, and to metabolize it into ATP, only the FLE synthesizes cysteine, threonine, tyrosine, tryptophan, phenylalanine, and serine from pyruvate, and can metabolize glutamate, glutamine, and asparagine into ATP. Of these amino acids, FLE most crucially provides a reliable source of the amino acid cysteine, which is found in very low concentrations in bovine blood<sup>1</sup>. FLE-Am can also synthesize glutamine from glutamic acid and ammonia, thus recycling cellular ammonia waste to useful products. In sum, our data indicate that the superior biosynthetic capability of FLE-Am confers a selective advantage, which could have led to FLE-Am recently replacing an ancestral symbiont (e.g. CLE) with reduced metabolic capacity in *A. maculatum*.

Symbiosis with bacteria allows arthropods to thrive on nutrient-poor diets<sup>2</sup>. However, the absolute reliance on a symbiont could eventually become disadvantageous to the host because massive gene loss and accumulation of mildly deleterious mutations will corrode the symbiont's metabolic capability<sup>7,9</sup>. One solution to escaping this symbiosis "rabbit hole" is to supplement or replace the ancient symbiont with new bacteria acquired from the environment<sup>9</sup>. This process has occurred in several lineages of insects (e.g. ref. 17). In *A. maculatum*, FLE-Am seems to have recently replaced an older endosymbiont, whose identity is unknown, but based on the current bacterial prevalence data it was most probably a CLE or a *Rickettsia*<sup>12</sup>. FLE-Am likely arose by the domestication of a mammalian pathogen that was vectored by the tick. This process could occur rapidly because the bacterium doesn't have to learn anew how to circumvent the tick's immune response, and there is no need to attenuate bacterial virulence towards the tick<sup>18</sup>. Additionally, because genes that promote pathogenesis through amino acid scavenging (e.g., FTT 0968c, *xasA*), macrophage survival (e.g., *carA*, *carB*, *bioF*), and intracellular replication (*aroE1*, *purMCD*, *purL*, and *purF*)<sup>13,14</sup> are retained in FLE-Am (Fig. 1), they could be key to its endosymbiotic lifestyle. Further evolutionary, functional and genomic studies of FLEs and CLEs from a wide array of soft and hard ticks will help us to better understand how pathogenic *Francisella* evolved into endosymbionts that supports the blood-dependent lifestyle of ticks, and how tick-associated *Coxiella* evolved into pathogenic *C. burnetii*<sup>12</sup> (Fig. 4).

## Methods

**Sequencing and bacterial identification.** DNA was extracted from a female *A. maculatum* procured from Oklahoma State University Tick Rearing Facility, as described previously<sup>3</sup>, and was sequenced using





**Figure 4. Evolution of mammalian pathogens and tick symbionts.** Mammalian pathogens could evolve from tick-associated bacteria by acquiring virulence genes (e.g., *Coxiella burnetii*). Conversely, evolution of tick endosymbionts from mammalian pathogens is associated with loss of virulence genes (e.g., FLE-Am).

Illumina Hi-Seq 2500 (100 cycles, paired-end) at OHSU MPSSR, yielding approximately 180 million read pairs. Low confidence reads were removed, and the identity of bacteria present in the tick microbiome were determined utilizing  $\geq 1$  kb contigs binned using MEGAN<sup>19</sup>.

**Genome Assembly.** Trimmed reads were assembled into contigs using IDBA-UD<sup>20</sup>, and by comparing them to *Francisella* genome sequences, FLE contigs were identified. All trimmed reads were mapped to these contigs and then reassembled with IDBA-UD and IDBA-hybrid into a final set of seven contigs (Supplementary Table S3), which were submitted to NCBI (Accession: LVCE000000000). The completeness of the assembled genome was examined using a single-copy gene database<sup>21</sup>, and as a control, an identical single-copy gene analysis was performed on *F. tularensis* subsp. *tularensis* SCHU S4 genome (NC\_006570.2) (Table 1).

**Genome Annotation.** FLE-Am contigs were annotated using NCBI Prokaryotic Genome Annotation Pipeline, and pseudogenes were verified manually. Protein-coding genes were binned into categories based on their role in primary metabolism, amino acid and nucleic acid synthesis, or vitamin and cofactor metabolism, and then subsequently compared to CLEAA to illustrate differences in their metabolic capabilities. A list of genes critical to the pathogenicity of *F. tularensis*<sup>13,14</sup> was used to identify both functional and pseudogenized versions of virulence genes present in FLE-Am (Supplementary Table S2). Intact protein-coding genes present in FLE-Am but not in *F. tularensis* SCHU-S4 are provided in Supplementary Table S4.

**Phylogenetic Analysis.** In addition to FLE-Am, we included all 44 fully sequenced *Francisella* genomes to generate a robust phylogenetic tree. Using reciprocal BLASTP, a subset of 442 orthologous genes (Supplementary Dataset 2) that were conserved in all genomes was identified. Nucleotide sequences were aligned using Clustal Omega<sup>22</sup>, ambiguously aligned regions were removed using Gblocks<sup>23</sup>, and jModelTest2 was used to select the GTR+I+G (General Time Reversible plus Invariant sites plus Gamma distribution) evolution model<sup>24</sup>. Using concatenated sequences, Maximum Likelihood trees with 1,000 bootstrap replicates were constructed using RAXML<sup>25</sup>. Bayesian trees with a chain length of 500,000 and a burn-in fraction of 25% and sampling every 100 trees were constructed using MrBayes<sup>26</sup>.

## References

1. Duarte, R. T., Carvalho Simões, M. C. & Sgarbieri, V. C. Bovine blood components: fractionation, composition, and nutritive value. *J. Agri. c Food. Chem.* **47**, 231–236 (1999).
2. Baumann, P. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annu. Rev. Microbiol.* **59**, 155–189 (2005).
3. Smith, T. A., Driscoll, T., Gillespie, J. J. & Raghavan, R. A *Coxiella*-like endosymbiont is a potential vitamin source for the lone star tick. *Genome Biol. Evol.* **7**, 831–838 (2015).
4. Ahantarig, A., Trinachartvanit, W., Baimai, V. & Grubhoffer, L. Hard ticks and their bacterial endosymbionts (or would be pathogens). *Folia Microbiol.* **58**, 419–428 (2013).
5. Petersen, J. M., Mead, P. S. & Schriefer, M. E. *Francisella tularensis*: an arthropod-borne pathogen. *Vet. Res.* **40**, 07 (2009).
6. Duron, O., Sidi-Boumedine, K., Rousset, E., Moutailler, S. & Jourdain, E. The Importance of Ticks in Q Fever Transmission: What Has (and Has Not) Been Demonstrated? *Trends Parasitol.* **31**, 536–552 (2015).
7. McCutcheon, J. P. & Moran, N. A. Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**, 13–26 (2012).

8. Budachetri, K. *et al.* An insight into the microbiome of the *Amblyomma maculatum* (Acari: Ixodidae). *J. Med. Entomol.* **51**, 119–129 (2014).
9. Bennett, G. M. & Moran, N. A. Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. *Proc. Natl. Acad. Sci. USA* **112**, 10169–10176 (2015).
10. Baldrige, G. D., Scoles, G. A., Burkhardt, N. Y., Schloeder, B., Kurtti, T. J. & Munderloh, U. G. Transovarial transmission of *Francisella*-like endosymbionts and *Anaplasma phagocytophilum* variants in *Dermacentor albipictus* (Acari: Ixodidae). *J. Med. Entomol.* **46**, 625–632 (2009).
11. Sjödin, A. *et al.* Genome characterisation of the genus *Francisella* reveals insight into similar evolutionary paths in pathogens of mammals and fish. *BMC Genomics* **13**, 268 (2012).
12. Duron, O. *et al.* The recent evolution of a maternally-inherited endosymbiont of ticks led to the emergence of the Q fever pathogen, *Coxiella burnetii*. *Plos Pathog* **11**, e1004892 (2015).
13. Rowe, H. M. & Huntley, J. E. From the outside-in: the *Francisella tularensis* envelope and virulence. *Front. Cell Infect. Microbiol.* **5**, 94 (2015).
14. Melhorn, K. L. & Charbit, A. The unraveling panoply of *Francisella tularensis* virulence attributes. *Curr. Opin. Microbiol.* **13**, 11–17 (2010).
15. Klyachko, O., Stein, B. D., Grindle, N., Clay, K. & Faqua, C. Localization and visualization of a *Coxiella*-type symbiont within the lone star tick, *Amblyomma americanum*. *Appl. Environ. Microbiol.* **73**, 6584–6594 (2007).
16. Wright, C. L., Sonenshine, D. E., Gaff, H. D. & Hynes, W. L. *Rickettsia parkeri* transmission to *Amblyomma americanum* by co-feeding with *Amblyomma maculatum* (Acari: Ixodidae) and potential for spillover. *J. Med. Entomol.* **52**, 1090–1095 (2015).
17. Koga, R. & Moran, N. A. Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont. *ISME J.* **8**, 1237–1246 (2014).
18. Clayton, A. L. *et al.* A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect-bacterial symbioses. *Plos Genet.* **8**, e1002990 (2012).
19. Huson, D. H., Auch, A. E., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386 (2007).
20. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).
21. Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K. L., Tyson, G. W. & Nielsen, P. H. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* **31**, 533–538 (2013).
22. Stevers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539–539 (2014).
23. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).
24. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9**, 772 (2012).
25. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
26. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).

## Acknowledgements

We thank Colleen Campbell, Jess Millar, Todd Smith, Marie Baker and Jim Archuleta for technical assistance. This work was supported by Portland State University.

## Author Contributions

J.G.G., A.S.M. and R.R. conducted the experiments, analyzed the data, and prepared the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Gerhart, J. G. *et al.* A *Francisella*-like endosymbiont in the Gulf Coast tick evolved from a mammalian pathogen. *Sci. Rep.* **6**, 33670; doi: 10.1038/srep33670 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016

**Manuscript in preparation:**

**Multiple acquisitions of *Francisella*-like endosymbionts by soft ticks**

Jonathan G. Gerhart<sup>1</sup>, Abraham S. Moses<sup>1</sup>, Libor Grubhoffer<sup>2</sup>, Rahul Raghavan<sup>1\*</sup>

<sup>1</sup>Department of Biology and Center for Life in Extreme Environments,  
Portland State University, Portland, Oregon, 97201, USA.

<sup>2</sup>Biology Centre, Institute of Parasitology AS CR, Department of Molecular  
Ecology of Vectors and Pathogens, Ceske Budejovice 37005, Czech Republic.

\*Author for correspondence:

Rahul Raghavan | Phone: 503-725-3872 | Email: rahul.raghavan@pdx.edu

## ABSTRACT

Endosymbiosis in arthropods involves intracellular bacteria that supply an array of benefits to the host. In ticks, endosymbionts likely supply essential amino acids and cofactors, and recycle ammonia into amino acids. Endosymbionts of ticks are of special interest due to their close evolutionary relationships with tick-vectored pathogens that impact livestock and human health. For instance, ticks contain *Coxiella*-like endosymbionts (CLEs) or *Francisella*-like endosymbionts (FLEs) that are closely related to human pathogens *Coxiella burnetii* and *Francisella tularensis*, respectively. We recently sequenced the genome of the FLE present in the hard tick *Amblyomma maculatum* (FLE-Am) and showed that it evolved from a pathogenic ancestor. In order to expand our understanding of the evolutionary histories of FLEs in ticks, in this study, we sequenced and analyzed the genome of the FLE in the soft tick *Ornithodoros moubata* (FLE-Om). While this project was in progress, the genome of an FLE in another soft tick *Argas (Persicargas) arboreus* was published. Utilizing the three genomes, we establish that all FLEs evolved from a common pathogenic ancestor, a relationship that is converse to that observed in *Coxiella* where the human pathogen likely evolved from a tick-associated non-pathogenic ancestor. Additionally, we show that the

FLEs in the two soft ticks were likely acquired independently, probably via horizontal transfer between arthropod hosts or from environmental sources.

## INTRODUCTION

Bacterial endosymbionts allow arthropods to subsist on unbalanced diets. For example, *Buchnera aphidicola* provides amino acids missing in the diet of aphids that feed exclusively on phloem (Baumann 2005; Bennett & Moran 2015). Similarly, ticks that depend solely on blood, a poor source for several amino acids and cofactors, contain *Coxiella*-like or *Francisella*-like endosymbionts (referred to as CLEs and FLEs, respectively) (Smith et al. 2015; Gottlieb et al. 2015; Gerhart et al. 2016). Ticks also vector mammalian pathogens *Coxiella burnetii* and *Francisella tularensis* that are closely related to the endosymbionts (Ahantarig et al. 2013; Petersen et al. 2009; Duron et al. 2015). The evolutionary relationships between endosymbionts and pathogens carried by ticks are not completely understood. Previous studies indicate that pathogen-symbiont transitions could occur in either direction: the mammalian pathogen *C. burnetii* probably originated from a tick-associated non-pathogenic ancestor (Duron et al. 2015; Moses et al. 2017), whereas the FLE in the hard tick *Amblyomma maculatum* (FLE-Am) likely evolved from a mammalian pathogen (Gerhart et al. 2016).

In this study, to better understand the evolution of FLEs in ticks, we sequenced the genome of an FLE found in the soft tick *Ornithodoros moubata* (FLE-Om), and compared it to the genome sequences of FLE-Am and *Francisella*

*persica* – an FLE present in the soft tick *Argas (Persicargas) arboreus* (previously referred to as *Argas persicus*) (Hinrichs et al. 2016). Our analyses show that all FLEs evolved from a common pathogenic ancestor, but the FLEs in the two soft ticks were likely acquired independently.

## RESULTS AND DISCUSSION

### FLE-Om has a reduced genome

Intracellular “coccoid” bacteria were described in the malphigian tubules and ovaries of *O. moubata* more than four decades back using electron microscopy (Reinhardt et al. 1972). Recently, 16S rDNA PCR was used to confirm that the endosymbiotic bacterium was closely related to the human pathogen *F. tularensis* (Noda et al. 1997). Using high-throughput sequencing, we analyzed genomic DNA isolated from an adult female *O. moubata* obtained from a long-term laboratory colony (Rego et al. 2005). Out of the ~180 million sequencing reads, a vast majority was of host origin, with ~15 million reads contributing to assembling the nine FLE-Om contigs ([Table S1](#)). In addition, low amounts of *Clostridium* spp. and *Burkholderia* spp. were also present in the tick (<0.5% of reads). No other bacterium was detected at significant levels, including the *Coxiella*-like “Symbiont A” described previously from another *O. moubata*

laboratory colony (Noda et al. 1997). We screened several male and female *O. moubata* adults using both bacterial and *Coxiella*-specific 16S rDNA PCR primers to confirm that FLE-Om but not the *Coxiella*-like “Symbiont A” has been retained in the *O. moubata* laboratory colony (Klyachko et al. 2007; Klindworth et al. 2013). As shown previously in insects, this observation suggests that while FLE-Om is likely a primary endosymbiont that provides a critical function to *O. moubata*, the *Coxiella*-like “symbiont A” is likely a secondary symbiont or a transient bacterium (Hall et al. 2016).

The ~1.57 Mb genome of FLE-Om is similar in size to that of other FLEs (FLE-Am and *F. persica*) but substantially smaller than the ~1.90 Mb genome of the human pathogen *F. tularensis* (Table 1). Similarly, the G+C content and the number of protein-coding genes are also considerably lower in the FLEs in comparison to *F. tularensis*, signifying a more intimate association between FLEs and their hosts, which leads to smaller, A+T-biased genomes (McCutcheon & Moran 2010). Intriguingly, FLE-Am and FLE-Om have significantly higher number of pseudogenes, and contain lower number of rRNA operons than *F. persica*, indicating that *F. persica*’s evolutionary history is different from that of the other two FLEs.



**All FLEs evolved from a common pathogenic ancestor, but *F. persica* and FLE-Om were acquired independently**

We estimated the phylogenetic positions of FLEs using 404 orthologous genes in 49 fully sequenced *Francisella* genomes ([Supplemental Dataset 1](#)). Our data shows that all FLEs originated from a common potentially pathogenic ancestor ([Figure 1A](#)). However, FLE-Om is more closely related to FLE-Am than to *F. persica*, although the hosts of FLE-Om and *F. persica* are more closely related to each other (both are soft ticks) ([Figure 1B](#)). This incongruence between tick phylogeny and FLE phylogeny indicate that the FLEs in the two soft ticks are likely of independent origin. This observation is not surprising because unlike matching host and symbiont phylogenetic trees observed in many insects (e.g. Chen et al. 1999; Bennett & Moran 2015; Jousset et al. 2009), host-endosymbiont coevolution is rare across ticks, probably due to horizontal transfer of symbionts between unrelated tick species (Duron et al. 2015). A recent PCR survey of 81 species of hard and soft ticks showed that FLEs in several soft ticks were more closely related to FLEs in hard ticks than to each other, denoting that FLEs were acquired multiple times by soft ticks (Duron et al. 2017). Based on these observations, it is likely that the two FLEs in soft ticks were acquired recently from independent sources: FLE-Om likely from a hard tick, and *F. persica* from an

as yet unknown (environmental) source. Horizontal transfer of endosymbionts are probably more prevalent among ticks than among insects because tick endosymbionts are not restricted to specialized host cells such as bacteriocytes, but are usually present in several tissues including in the salivary glands (Klyachko et al. 2007; Budachetri et al. 2014). The presence of FLEs in the salivary glands of ticks could be a holdover from their previous pathogenic lifestyle, and it could facilitate horizontal transfer between blood-feeding arthropods while co-feeding on the same vertebrate host (Wright et al. 2015; Lee et al. 2016).

### **Metabolic and virulence features of FLEs**

The FLE pan-genome consists of 1395 genes, with the majority (751) present in the core genome ([Figure 2](#)). We examined the metabolic potential of FLEs ([Figure 3](#)) and found that *F. persica* encodes a large, unique biosynthetic pathway for nicotinamide, while lacking complete pathways for pantothenate and heme, which are present in FLE-Om and FLE-Am. FLE-Am produces the most diverse group of amino acids, while FLE-Om has no unique pathways. All endosymbionts have a TCA cycle, electron transport chain, a glutamine/glutamate synthesis pathway that could recycle waste ammonia into amino acids, and cysteine, cystathione and glycine conversion enzymes. Lastly,

within the auxiliary genomes, serine and chorismate synthesis pathways are shared between FLE-Om and FLE-Am, whereas glutamate semialdehyde dehydrogenase and asparagine deaminase are shared between FLE-Am and *F. persica*. In sum, the FLEs have the potential to synthesize several amino acids and cofactors present in very low concentrations in their hosts' blood meal ([Table S2](#)), indicating that they could improve host fitness by functioning as nutrient provisioning endosymbionts. Furthermore, the extensive metabolic capabilities of FLEs could provide them a selective advantage over endosymbionts with limited metabolic potential such as ancestral CLEs with degraded genomes, thereby resulting in FLEs replacing older endosymbionts (Bennett & Moran 2015; Gerhart et al. 2016; Duron et al. 2017). Lastly, unlike for most other arthropod endosymbionts, the broad biosynthetic capability of *F. persica* has allowed its laboratory culture (Hinrichs et al. 2016), and it should be possible to develop axenic growth media for FLE-Am and FLE-Om using our genomic data, as previously done for other obligate intracellular bacteria (Omsland et al. 2009).

Reflective of their pathogenic heritage, all three FLEs encode several genes required for virulence in pathogenic *Francisella* (Rowe & Huntley 2015; Meibom & Charbit 2010). As expected in nutrient provisioning endosymbionts, most virulence genes have been inactivated or deleted in FLE-Am and FLE-Om ([Table](#)

2, Table S3). *F. persica*, however, contains considerably higher number of potentially functional virulence genes, including in important structures such as LPS and Type VI secretion system, suggesting that it is still pathogenic to mammals. In fact, *F. persica* (previously misidentified as *Wolbachia persica*) was found to cause fever and death in guinea pigs (Suitor & Weiss 1961). Intriguingly, MA067296, a strain of *Francisella* that caused an opportunistic human infection (Kugeler et al. 2008; Challacombe et al. 2017), is more closely related to the FLEs than to the human pathogen *F. tularensis* (Figure 1A), supporting our conclusion that the ancestor of FLEs was a mammalian pathogen.

In conclusion, we show that while all three FLEs originated from a common pathogenic ancestor and could function as nutrient-provisioning endosymbionts, the two FLEs in soft ticks have divergent evolutionary histories and dissimilar virulence potentials, indicative of independent acquisition events, probably of preexisting endosymbionts or from environmental sources (Clayton et al. 2012; Duron et al. 2017). Further studies are required to understand the functions of FLEs in ticks, to delineate the evolutionary histories of all FLEs in hard and soft ticks, and to define the possible relationship between FLEs and environmental *Francisella* (Barns et al. 2005; Keim et al. 2007).

## **MATERIALS AND METHODS**

### **Sequencing and bacterial identification.**

A female *O. moubata* from a laboratory colony maintained at the Institute of Parasitology, Czech Republic was used in this study. Tick outer surface was sterilized using 70% ethanol, and midgut tissue was extracted and subjected to Protease K digestion, followed by DNA extraction using a DNeasy kit and protocol (Qiagen). Purified DNA was sequenced in a single lane of Illumina Hi-Seq 2500 (100 bp, paired-end) at Oregon Health and Science University's Massively Parallel Sequencing Shared Resource, yielding approximately 180 million read pairs. Low confidence reads were removed using Trimmomatic (Bolger et al. 2014). The identity of bacteria present in the tick microbiome were determined after the first round of de novo assembly with IDBA (Peng et al. 2012) utilizing  $\geq 1$  kb contigs identified via Blastn and binned according to the genus of closet hit using MEGAN5's taxonomy browser (Huson et al. 2007).

### **Genome assembly.**

Trimmed reads were assembled into contigs using IDBA, using a k-mer range of 49:121, a step size of two, and a minimum contig length of 1 kb. Putative FLE contigs were identified through Blastn against a library of *Francisella* genome

sequences obtained from NCBI. All reads were mapped to these contigs using bowtie2 and Samtools (Langmead & Salzberg 2012), and the mapped reads along with 20% of original reads were reassembled into a final set of nine contigs, which have been submitted to NCBI (accession: LVCE000000000). The completeness of the assembled genome was examined using a single-copy gene database (Albertsen et al. 2013), yielding identical results for FLE-Om, FLE-Am, *Francisella persica*, and *Francisella tularensis* (106/111 genes).

#### **Genome annotation and assessment of metabolic and virulence potential.**

FLE-Om contigs were annotated using NCBI's prokaryotic genome annotation pipeline, and pseudogenes were verified manually. Protein-coding genes were binned into categories based on their roles in primary metabolism, amino acid and nucleic acid synthesis, or vitamin and cofactor metabolism using KEGG (Manyam et al. 2015) and MEGAN5. Genes present in FLE-Am, FLE-Om and *F. persica* were compared to illustrate similarities and differences in their metabolic capabilities. A list of genes critical to the pathogenicity of *F. tularensis* was used to identify functional and pseudogenized versions of virulence genes present in FLE-Om, FLE-Am and *F. persica* (Rowe & Huntley 2015; Meibom & Charbit 2010). For metabolic reconstruction, groups of shared and unique genes were

separately loaded into MEGAN5's KEGG mapping function and the completeness of metabolic pathways was inferred. If a few genes within a larger pathway were missing from the original analysis, it was confirmed or corrected using Blast searches.

### **Pan-genome construction.**

The pan-genome of FLEs was constructed using a Blastn database created from the genes present in all three genomes using a criteria of at least 80% nucleotide identity over a region of at least 90% to be considered homologous. Biochemical, metabolic and cellular functions were assigned to each group by utilizing KEGG pathways viewer in MEGAN5. Diagrams and figures were generated using Circos (Krzywinski et al. 2009), and Adobe Illustrator.

### **Phylogenetic analysis.**

In addition to FLE-Om, FLE-Am and *F. persica*, we included 46 fully sequenced *Francisella* genomes to generate a robust phylogenetic tree (Gerhart et al. 2016). Using reciprocal Blastp, a subset of 404 orthologous genes were identified. Sequences were aligned using Clustal Omega (Sievers et al. 2014) and trimmed using Gblocks (Talavera & Castresana 2007). jModelTest2 was used to select the

GTR+I+G (General Time Reversible plus Invariant sites plus Gamma distribution) evolution model (Darriba et al. 2012). Maximum Likelihood and Bayesian trees were constructed using RAxML (Stamatakis 2014) and MrBayes (Huelsenbeck & Ronquist 2001), respectively.

## **ACKNOWLEDGEMENTS**

We thank Colleen Campbell, Jess Millar, Todd Smith and Jim Archuleta for technical assistance. This work was supported in part by National Institutes of Health grants AI123464 and AI126385 to RR.

## **REFERENCES**

- Ahantarig A, Trinachartvanit W, Baimai V, Grubhoffer L. 2013. Hard ticks and their bacterial endosymbionts (or would be pathogens). *Folia Microbiol. (Praha)*. 58:419–428. doi: 10.1007/s12223-013-0222-1.
- Albertsen M et al. 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* 31:533–538. doi: 10.1038/nbt.2579.
- Barns SM, Grow CC, Okinaka RT, Keim P, Kuske CR. 2005. Detection of diverse



- new *Francisella*-like bacteria in environmental samples. Appl. Environ. Microbiol. 71:5494–5500. doi: 10.1128/AEM.71.9.5494-5500.2005.
- Baumann P. 2005. Biology bacteriocyte-associated endosymbionts of plant sap-sucking insects. Annu. Rev. Microbiol. 59:155–189. doi: 10.1146/annurev.micro.59.030804.121041.
- Bennett GM, Moran NA. 2015. Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. Proc. Natl. Acad. Sci. USA. 112:10169–10176. doi: 10.1073/pnas.1421388112.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120. doi: 10.1093/bioinformatics/btu170.
- Budachetri K et al. 2014. An insight into the microbiome of the *Amblyomma maculatum* (Acari: Ixodidae). J. Med. Entomol. 51:119–29. doi: 10.1603/ME12223.
- Challacombe JF et al. 2017. Whole-genome relationships among *Francisella* bacteria of diverse origins define new species and provide specific regions for detection. Appl. Environ. Microbiol. 83:e02589–16. doi: 10.1128/AEM.02589-16.
- Chen X, Li S, Aksoy S. 1999. Concordant evolution of a symbiont with its host

insect species: molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. J. Mol. Evol. 48:49–58. doi: 10.1007/PL00006444.

Clayton AL et al. 2012. A Novel Human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect–bacterial symbioses. PLoS Genet. 8:e1002990. doi: 10.1371/journal.pgen.1002990.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods. 9:772–772. doi: 10.1038/nmeth.2109.

Duron O et al. 2017. Evolutionary changes in symbiont community structure in ticks. Mol. Ecol. 38:42–49. doi: 10.1111/mec.14094.

Duron O et al. 2015. The recent evolution of a maternally-inherited endosymbiont of ticks led to the emergence of the Q fever pathogen, *Coxiella burnetii*. PLOS Pathog. 11:e1004892. doi: 10.1371/journal.ppat.1004892.

Gerhart JG, Moses AS, Raghavan R. 2016. A *Francisella*-like endosymbiont in the Gulf Coast tick evolved from a mammalian pathogen. Sci. Rep. 6:33670. doi: 10.1038/srep33670.

Gottlieb Y, Lalzar I, Klasson L. 2015. Distinctive genome reduction rates revealed

- by genomic analyses of two *Coxiella*-like endosymbionts in ticks. *Genome Biol. Evol.* 7:1779–1796. doi: 10.1093/gbe/evv108.
- Hall AAG et al. 2016. Codivergence of the primary bacterial endosymbiont of psyllids versus host switches and replacement of their secondary bacterial endosymbionts. *Environ. Microbiol.* 18:2591–2603. doi: 10.1111/1462-2920.13351.
- Hinrichs SH et al. 2016. Reclassification of *Wolbachia persica* as *Francisella persica* comb. nov. and emended description of the family Francisellaceae. *Int. J. Syst. Evol. Microbiol.* 66:1200–1205. doi: 10.1099/ijsem.0.000855.
- Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics.* 17:754–755. doi: 10.1093/bioinformatics/17.8.754.
- Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. *Genome Res.* 17:377–386. doi: 10.1101/gr.5969107.
- Jousselin E, Desdevises Y, Coeur d’acier A. 2009. Fine-scale cospeciation between *Brachycaudus* and *Buchnera aphidicola*: bacterial genome helps define species and evolutionary relationships in aphids. *Proc. R. Soc. B Biol. Sci.* 276:187–196. doi: 10.1098/rspb.2008.0679.
- Keim P, Johansson A, Wagner DM. 2007. Molecular epidemiology, evolution, and ecology of *Francisella*. *Ann. N. Y. Acad. Sci.* 1105:30–66. doi:

10.1196/annals.1409.011.

Klindworth A et al. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41:e1–e1. doi: 10.1093/nar/gks808.

Klyachko O, Stein BD, Grindle N, Clay K, Fuqua C. 2007. Localization and visualization of a *Coxiella*-type symbiont within the lone star tick, *Amblyomma americanum*. *Appl. Environ. Microbiol.* 73:6584–6594. doi: 10.1128/AEM.00537-07.

Krzywinski M et al. 2009. Circos: An information aesthetic for comparative genomics. *Genome Res.* 19:1639–1645. doi: 10.1101/gr.092759.109.

Kugeler KJ et al. 2008. Isolation and characterization of a novel *Francisella* sp. from human cerebrospinal fluid and blood. *J. Clin. Microbiol.* 46:2428–2431. doi: 10.1128/JCM.00698-08.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods.* 9:357–359. doi: 10.1038/nmeth.1923.

Lee S-H et al. 2016. Novel Detection of *Coxiella* spp., *Theileria luwenshuni*, and *T. ovis* endosymbionts in deer keds (*Lipoptena fortisetosa*). *PLoS One.* 11:e0156727. doi: 10.1371/journal.pone.0156727.

Manyam G, Birerdinc A, Baranova A. 2015. KPP: KEGG pathway painter. *BMC*

- Syst. Biol. 9:S3. doi: 10.1186/1752-0509-9-S2-S3.
- McCutcheon JP, Moran NA. 2010. Functional convergence in reduced genomes of bacterial symbionts spanning 200 million years of evolution. *Genome Biol. Evol.* doi: 10.1093/gbe/evq055.
- Meibom KL, Charbit A. 2010. The unraveling panoply of *Francisella tularensis* virulence attributes. *Curr. Opin. Microbiol.* 13:11–17. doi: 10.1016/j.mib.2009.11.007.
- Moses AS, Millar JA, Bonazzi M, Beare PA, Raghavan R. 2017. Horizontally acquired biosynthesis genes boost *Coxiella burnetii*'s physiology. *Front. Cell. Infect. Microbiol.* 7. doi: 10.3389/fcimb.2017.00174.
- Noda H, Munderloh UG, Kurtti TJ. 1997. Endosymbionts of ticks and their relationship to *Wolbachia* spp. and tick-borne pathogens of humans and animals. *Appl. Environ. Microbiol.* 63:3926–32. doi: 0099-2240/97/\$04.00+0.
- Omsland A et al. 2009. Host cell-free growth of the Q fever bacterium *Coxiella burnetii*. *Proc. Natl. Acad. Sci. USA.* 106:4430–4434. doi: 10.1073/pnas.0812074106.
- Peng Y, Leung HCM, Yiu SM, Chin FYL. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics.* 28:1420–1428. doi: 10.1093/bioinformatics/bts174.

- Petersen JM, Mead PS, Schrieffer ME. 2009. *Francisella tularensis* : an arthropod-borne pathogen. Vet. Res. 40:07. doi: 10.1051/vetres:2008045.
- Rego ROM et al. 2005. Molecular cloning and comparative analysis of fibrinogen-related proteins from the soft tick *Ornithodoros moubata* and the hard tick *Ixodes ricinus*. Insect Biochem. Mol. Biol. 35:991–1004. doi: 10.1016/j.ibmb.2005.04.001.
- Reinhardt C, Aeschlimann A, Hecker H. 1972. Distribution of *Rickettsia*-like microorganisms in various organs of an *Ornithodoros moubata* laboratory strain (Ixodoidea, Argasidae) as revealed by electron microscopy. Zeitschrift fur Parasitenkd. 39:201–209. doi: 10.1007/BF00329456.
- Rowe HM, Huntley JF. 2015. From the Outside-In: The *Francisella tularensis* envelope and virulence. Front. Cell. Infect. Microbiol. 5:1–20. doi: 10.3389/fcimb.2015.00094.
- Sievers F et al. 2014. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7:539–539. doi: 10.1038/msb.2011.75.
- Smith TA, Driscoll T, Gillespie JJ, Raghavan R. 2015. A *Coxiella*-like endosymbiont is a potential vitamin source for the lone star tick. Genome Biol. Evol. 7:831–838. doi: 10.1093/gbe/evv016.

- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30:1312–1313. doi: 10.1093/bioinformatics/btu033.
- Suitor EC, Weiss E. 1961. Isolation of a Rickettsialike microorganism (*Wolbachia persica*, N. SP.) from *Argas persicus* (Oken). *J. Infect. Dis.* 108:95–106. doi: 10.1093/infdis/108.1.95.
- Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56:564–577. doi: 10.1080/10635150701472164.
- Wright CL, Sonenshine DE, Gaff HD, Hynes WL. 2015. *Rickettsia parkeri* transmission to *Amblyomma americanum* by cofeeding with *Amblyomma maculatum* (Acari: Ixodidae) and potential for spillover. *J. Med. Entomol.* 52:1090–1095. doi: 10.1093/jme/tjv086.

## FIGURE LEGENDS

**Figure 1. FLEs evolved from a pathogenic ancestor but were acquired multiple times. (A)** Phylogenetic tree based on 404 orthologous genes in 49 fully sequenced genomes of *Francisella*. MA067296 is an opportunistic human pathogen, and TX077308 is an environmental strain isolated from seawater

(Challacombe et al. 2017). **(B)** FLEs phylogeny (left) does not correspond to the host phylogeny (right; based on an 18S rDNA sequences), indicating that FLEs did not diverge along with their hosts. Branch lengths are not representative of evolutionary distance. Soft ticks and their FLEs are highlighted in green boxes.

**Figure 2. Comparison of the three FLE genomes.** From the outside: (1) coding domains in both forward and reverse directions in blue, (2) pathogenic genes in orange and pseudogenized pathogenic genes in black, (3) identified ribosomal RNA genes, and (4) read coverage for FLE-Om and FLE-Am in green and blue, respectively. The inner Venn diagram depicts the FLE pan-genome, showing conserved, peripheral and unique genes in each bacterium.

**Figure 3. Comparison of metabolic features of FLEs.** Arrows indicate simple one- or two-enzyme processes. Larger, multi-step biosynthetic pathways within boxes.



**Table 1.** Genome features of *Francisella* species.

Feature	FLE-Om	FLE-Am	<i>F. persica</i>	<i>F. tularensis</i>
Contigs	9	7	1	1
Length (bp)	1,574,786	1,556,255	1,540,154	1,892,772
%G+C	31.80%	31.80%	31.39%	32.30%
Coding domains	992	1,001	1,096	1,556
Coding density	55.51%	56.90%	68%	92%
Average gene length	881 bp	886 bp	964 bp	937 bp
Pseudogenes	543	484	205	227
Single copy genes <sup>a</sup>	106/111	106/111	106/111	106/111
rRNAs	7	7	9	10
tRNAs	34	32	38	38
Accession #	LVCE00000000	LNCT000000	NZ_CP01302	NC_006570
	0	00	2	

<sup>a</sup> Albertsen et al. 2013

**Table 2.** Virulence genes in *Francisella* species.

	FLE-Om genes		FLE-Am genes		<i>F. persica</i> genes		<i>F. tularensis</i> genes	
	Codin g	Pseud o	Codin g	Pseud o	Codin g	Pseud o	Codin g	Pseud o
LPS and capsule synthesis	9	2	8	6	13	1	22	0
Antigen synthesis	5	3	6	1	5	2	11	1
Type 4 Pili	2	6	4	10	8	4	17	0
Hypothetical porins	1	0	1	3	2	0	2	0
Outer membrane proteins	5	0	6	2	8	0	8	0
Inner membrane proteins	8	1	8	0	8	0	10	0
Type VI Secretion System	1	6	3	4	16	0	17	0

Figures:

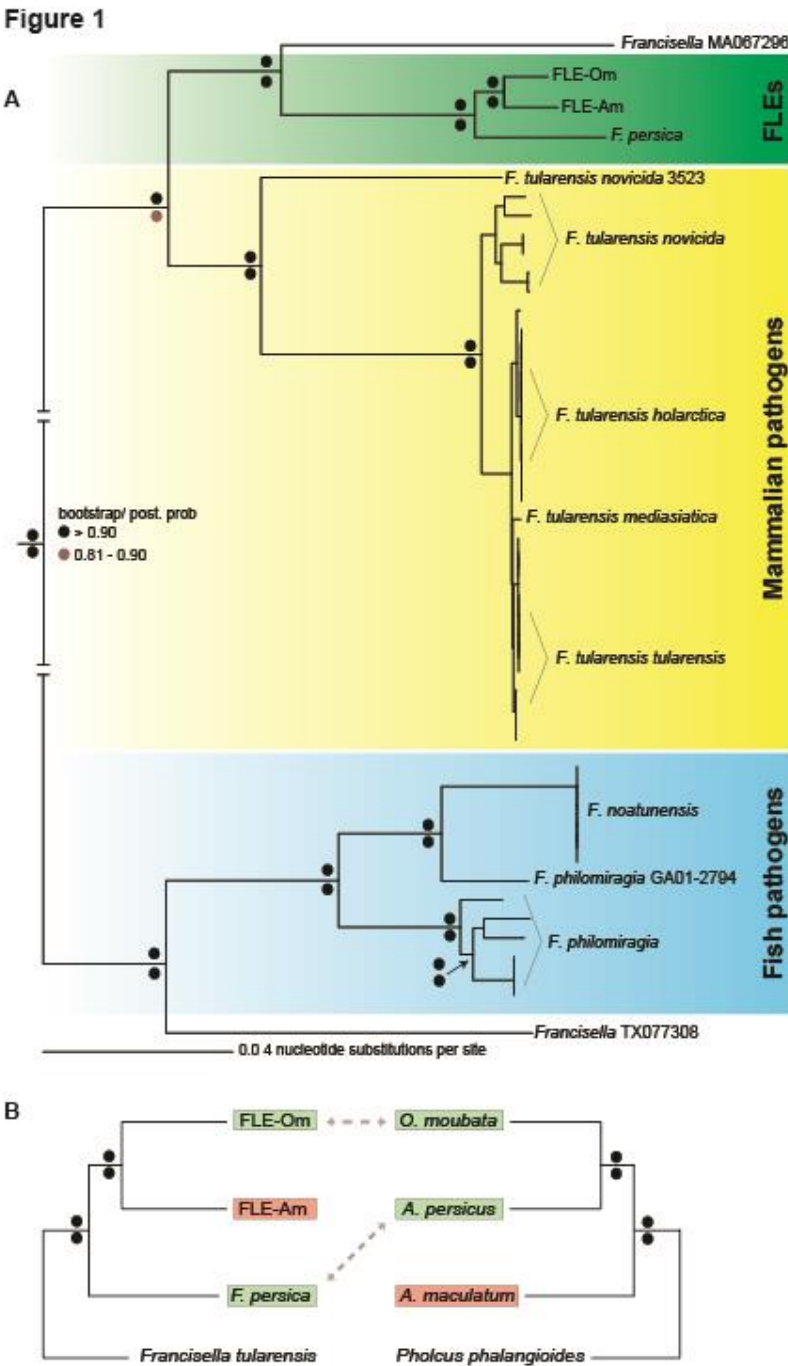


Figure 2

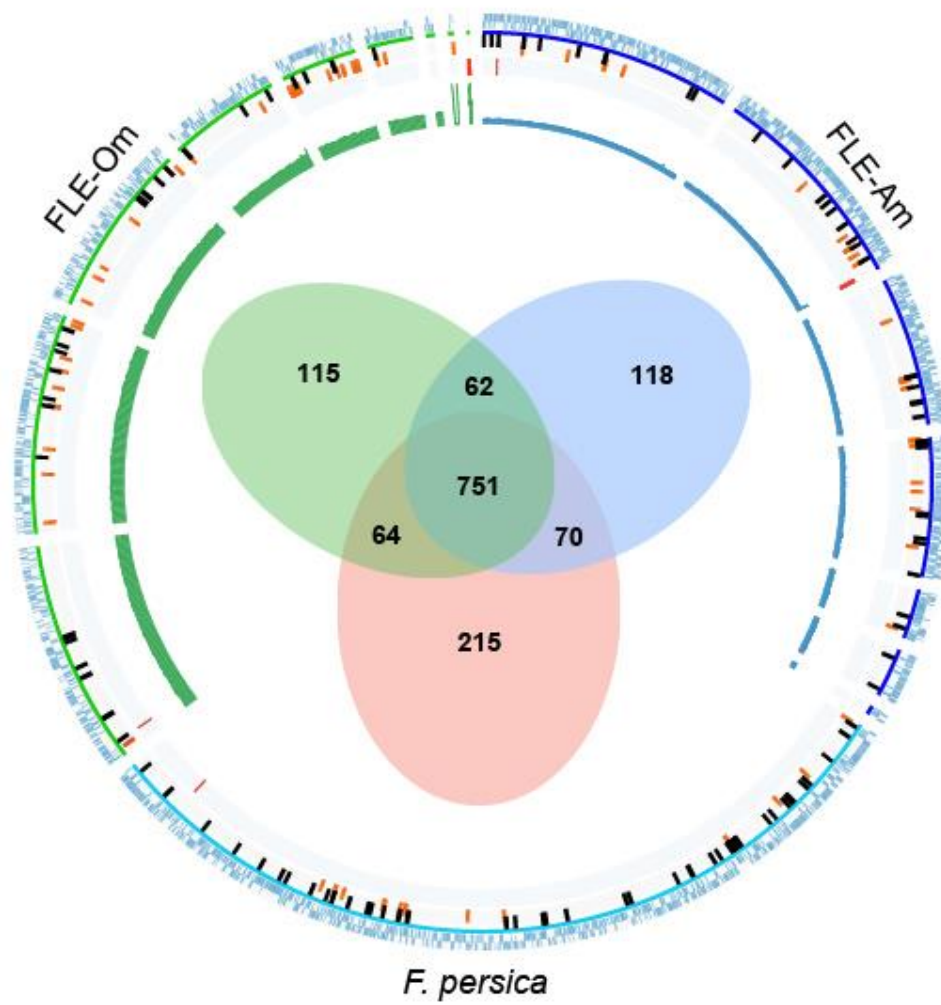
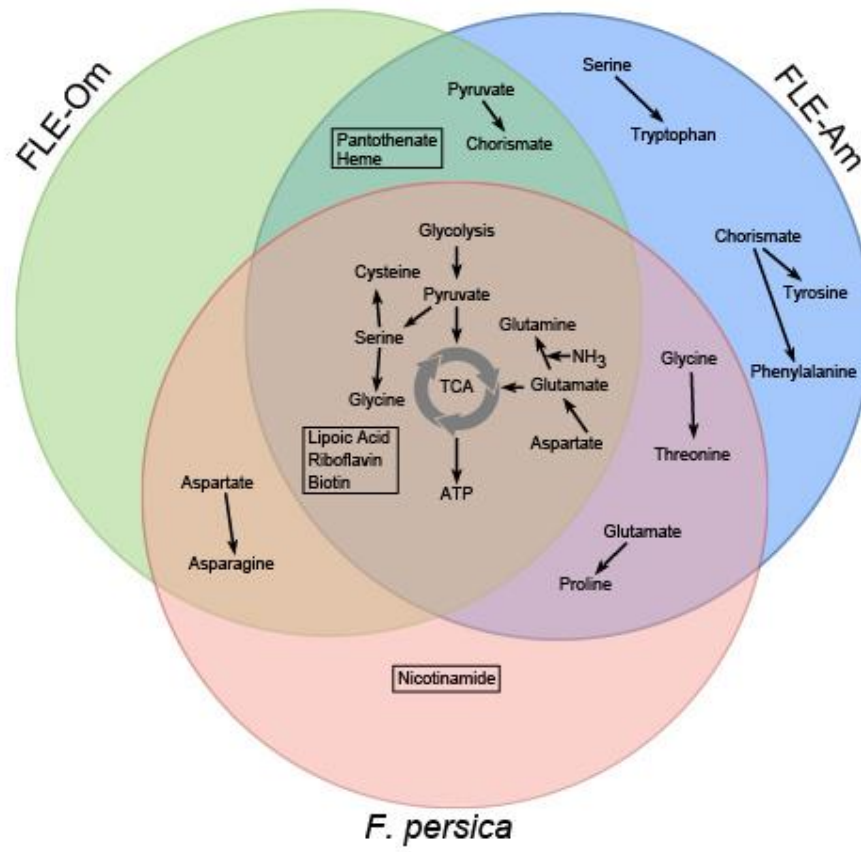


Figure 3



**Table S1.** Summary of FLE genome assemblies.

Contig ID	Length	Reads Mapped	Coverage Depth
<b>FLE-Am</b>			
Full Assembly	1,556,255	1,285,678	82.61
Scaffold 1	454,721	355,833	75.88
Scaffold 2	386,665	341,266	84.30
Scaffold 3	269,887	223,506	77.35
Scaffold 4	250,614	197,334	76.11
Scaffold 5	89,993	77,907	83.26
Scaffold 6	88,538	76,923	84.96
Scaffold 7	15,837	12,909	77.21
<b>FLE-Om</b>			
Full Assembly	1,574,777	14,652,882	930.47
Scaffold 1	407,560	3,077,041	734.15
Scaffold 2	395,818	3,021,047	741.58
Scaffold 3	315,083	2,255,793	695.16
Scaffold 4	203,407	1,472,195	698.97
Scaffold 5	138,216	1,026,455	720.35
Scaffold 6	83,179	631,584	738.22
Scaffold 7	16,010	120,852	730.51
Scaffold 8	10,589	2,953,267	20,521.68
Scaffold 9	4,915	94,648	1,693.71

**Table S2.** Amino acids and cofactors produced by FLEs.

Product	FLE-Am	FLE-Om	<i>F. persica</i>
Glutamine	+	+	+
Glutamate	+	+	+
Glycine	+	+	+
Serine	+	+	+
Asparagine	-	+	+
Alanine	+	+	-
Aspartate	+	+	+
Cysteine	+	+	+
Tryptophan	+	-	-
Threonine	+	-	+
Proline	+	-	+
Phenylalanine	+	-	-
Tyrosine	+	-	-
Riboflavin	+	+	+
FADH	+	+	+
Biotin	+	+	+
Lipoic acid	+	+	+
Folate	+	+	+
Heme	+	+	-
CoA	+	+	+
Nicotinamide	-	-	+
Pantothenate	+	+	-

**Table S3.** Virulence genes in *Francisella*.

Category <sup>a</sup>	Locus tag <sup>b</sup>	Function	<i>F. tularensis</i> <sup>c</sup>	<i>F. persica</i>	FLE-Am	FLE-Om
LPS and capsule synthesis	FTT 0286c	LPS	+	+	Φ	Φ
	FTT 0454	LPS	+	+	Φ	+
	FTT 0455c	LPS	+	+	+	+
	FTT 0789	Capsule	+	+	+	+
	FTT 0790	Capsule	+	+	Φ	Φ
	FTT 0791	Capsule	+	+	Φ	+
	FTT 0792	Capsule	+	-	-	-
	FTT 0793	Capsule	+	-	-	-
	FTT 0794	Capsule	+	-	-	-
	FTT 0795	Capsule	+	-	-	-
	FTT 0796	Capsule	+	-	-	-
	FTT 0797	Capsule	+	-	-	-
	FTT 0798	Capsule	+	-	-	-
	FTT 0799	Capsule	+	-	-	-
	FTT 0800	Capsule	+	+	Φ	Φ
	FTT 0805	Unknown	+	+	+	+
	FTT 0807	Unknown	+	+	+	+
	FTT 0806	Unknown	+	+	Φ	+
	FTT 0891	LPS	+	+	+	+
	FTT 1236	Capsule	+	+	+	+
	FTT 1237	LPS	+	+	+	Φ
	FTT 1238c	Capsule	+	+	+	-
	FTT 1450c	Both	+	+	Φ	+
	FTT 1453c	LPS	+	-	-	-
	FTT 1455c	Both	+	-	-	-
	FTT 1458c	LPS	+	-	-	-
	FTT 1460c	LPS	+	-	-	-
	FTT 1461c	LPS	+	-	-	-
	FTT 1462c	Both	+	+	+	-
	FTT 1463c	Both	+	+	+	+
	FTT 1464c	Both	+	+	+	+
	FTT 1561	LPS	+	+	+	+
	FTT 1571c	LPS	+	+	+	+



Type 4 pili	FTT 1643c	LPS	Φ	+	+	Φ
	FTT 1134	Pilin	+	+	Φ	Φ
	FTT 0890c	Pilin	+	+	Φ	Φ
	FTT 0889c	Pilin	+	Φ	Φ	Φ
	FTT 0888c	Pilin	+	-	Φ	-
	FTT 0861c	Pilin	+	Φ	Φ	Φ
	FTT 0230c	Pilin	+	+	+	-
	FTT 1341c	Pilin	+	+	+	+
	FTT 1133	Assembly ATPase	+	+	Φ	Φ
	FTT 0905	PilE1 glycolyase	+	Φ	Φ	Φ
	FTT 1156c	OM unit	+	+	Φ	Φ
	FTT 0088	Retraction ATPase	+	-	Φ	Φ
	FTT 0715	Secreted	+	-	-	-
	FTT 1786c	Secreted	+	-	-	-
	FTT 1577	Secreted	+	-	Φ	Φ
	FTN 1186	Secreted	-	-	-	-
	FTT 1069	Secreted	+	-	-	-
	FTT 0580	Secreted	+	+	+	+
	FTT 1330	Secreted	+	+	+	-
Hypothetical porins	FTT 0025c	Outer Membrane	+	+	ΦΦΦ	+
	FTT 0119	Outer Membrane	+	+	+	-
Outer membrane proteins	FTT 0369c	Outer Membrane	+	+	+	+
	FTT 0583	Outer Membrane	+	+	+	+
	FTT 0831c	Outer Membrane	+	+	+	+
	FTT 0918	Outer Membrane	+	+	+	+
	FTT 1103	Outer Membrane	+	+	+	+
	FTT 1346	Outer Membrane	+	-	Φ	Φ
	FTT 1416c	Outer Membrane	+	+	Φ	Φ
	FTT 1724c	Outer Membrane	+	+	+	Φ
Inner membrane proteins	FTT 0094c	Inner Membrane	+	+	+	-
	FTT 0181c	Inner Membrane	+	+	+	+
	FTT 0345	Inner Membrane	+	+	+	+

Secretion systems	FTT 0398c	Inner Membrane	+	+	+	+
	FTT 0399c	Inner Membrane	+	+	+	+
	FTT 0404	Inner Membrane	+	+	+	+
	FTT 0715	Inner Membrane	+	-	-	+
	FTT 0879	Inner Membrane	+	+	+	-
	FTT 0891	Periplasm	+	+	+	+
	FTT 1736c	Inner Membrane	+	-	-	+
	FTT 0068	Secreted	+	+	+	+
	FTT 0221	Secreted	+	-	-	-
	FTT 0484	Secreted	+	+	+	+
		Secreted Beta	+	Φ	Φ	Φ
	FTT 0611c	Lactamase		Φ	Φ	Φ
	FTT 0721	Secreted	+	+	Φ	Φ
	FTT 1095	Type 1 System	+	+	+	Φ
	FTT 1258	Type 1 System	+	+	-	Φ
	FTT 1269c	Secreted DNA K	+	+	+	+
	FTT 1344	Type 6 System	+	+	Φ	Φ
	FTT 1345	Type 6 System	+	+	Φ	Φ
	FTT 1346	Type 6 System	+	+	Φ	-
	FTT 1347	Type 6 System	+	+	+	Φ
	FTT 1348	Type 6 System	+	+	-	-
	FTT 1349	Type 6 System	+	+	+	-
	FTT 1350	Type 6 System	+	+	-	-
	FTT 1351	Type 6 System	+	+	-	Φ
	FTT 1352	Type 6 System	+	+	-	-
	FTT 1353	Type 6 System	+	+	-	Φ
	FTT 1354	Type 6 System	+	+	Φ	Φ
	FTT 1355	Type 6 System	+	+	+	Φ
	FTT 1356c	Type 6 System	+	+	-	-
	FTT 1357c	Type 6 System	+	+	-	-
	FTT 1358c	Type 6 System	+	+	-	-
	FTT 1359c	Type 6 System	+	+	-	-
	FTT 1360c	Type 6 System	+	-	-	-
	FTT 1361c	Secreted	Φ	+	+	+
	FTT 1441	Secreted	+	+	Φ	Φ
	FTT 1695	Secreted	+	+	+	+
	FTT 1696	Secreted	+	+	+	+
	FTT 1724c	Type 1 System	+	+	+	-

<sup>a</sup>Based on Meibom et al. 2010 and Rowe et al. 2015.

<sup>b</sup>From NC\_006570.2

<sup>c</sup>Gene present (+), absent (-), pseudogenized (Φ).